



**UNIVERSIDAD NACIONAL AUTÓNOMA DE  
MÉXICO**

---

---

**FACULTAD DE ESTUDIOS SUPERIORES  
ARAGÓN**

**EL ANÁLISIS DE BIG DATA COMO RECURSO DE  
INVESTIGACIÓN EN LA SOCIOLOGÍA**

**T E S I S**

**PARA OBTENER EL TÍTULO DE  
LICENCIADO EN SOCIOLOGÍA**

**P R E S E N T A**

**ISLAS MONROY EDILBERTO**

**ASESORA: MTRA. MARÍA ISABEL CHÁVEZ GARFIAS**



Ciudad Nezahualcóyotl, Estado de México, 2021



Universidad Nacional  
Autónoma de México



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

## **Agradecimientos**

A mis padres.

Por brindarme su guía, su apoyo y confianza en las decisiones que he tomado a lo largo del camino.

A mis amigos.

Que han sido ese escape necesario para despejar la mente y disfrutar de momentos muy amenos.

A Liz.

Por ser mi compañera de vida, mi mejor amiga y mi confidente, gracias por todo el apoyo que me has brindado durante estos años.

Ti voglio molto bene.

A la Mtra. Isabel Chávez.

Que pese a ser un tema nuevo para ella se aventuró a guiarme y darme su consejo en la elaboración de este trabajo.

# Índice

Introducción .....	2
<b>Capítulo 1. El almacenamiento digital y los sistemas de información como origen del Big Data.....</b>	<b>10</b>
1.1 El almacenamiento digital y los sistemas de información.....	10
1.2 Los Data warehouse y Data lake; el cauce de información para Big Data .....	16
1.3 ¿Qué es el Big Data?.....	19
1.4 Las características del Big Data.....	24
1.4.1 <i>Volumen</i> .....	24
1.4.2 <i>Variiedad o variabilidad</i> .....	25
1.4.3 <i>Velocidad</i> .....	26
1.4.4 <i>Veracidad</i> .....	27
1.4.5 <i>Valor</i> .....	27
1.4.6 <i>Otras dimensiones consideradas</i> .....	28
1.5 ¿Qué tipo de información se puede encontrar en Big Data? .....	30
<b>Capítulo 2. La importancia de analizar el Big Data .....</b>	<b>33</b>
2.1 El Big Data en la investigación.....	36
2.2 Los casos de uso del Big Data (casos prácticos del Big Data en el mundo real) 40	
2.3 Los niveles de profundidad de Big Data .....	48
2.3.1 <i>Open Data</i> .....	49
2.3.2 <i>Social Data</i> .....	55
2.3.3 <i>Geo Data</i> .....	66
<b>Capítulo 3. El Big Data desde una perspectiva sociológica .....</b>	<b>71</b>
3.1 Ecosistemas digitales para utilización del Big Data en la sociología .....	74
3.2 Recursos de la sociología en el Big Data.....	77
3.3 La analítica de Big Data desde una perspectiva sociológica .....	80
3.3.1 <i>Descriptivo</i> .....	83
3.3.2 <i>Diagnóstico</i> .....	85
3.3.3 <i>Predictivo</i> .....	87
3.3.4 <i>Prescriptivo</i> .....	91
<b>Conclusiones. Ventajas y desventajas del Big Data en la sociología.....</b>	<b>93</b>
Bibliografía.....	106

## Introducción

En las últimas dos décadas los avances tecnológicos han permitido que la sociedad tenga un acercamiento importante con la tecnología, a tal grado que actualmente, sin darnos cuenta, ya llevamos en nuestro bolsillo dispositivos que son capaces de realizar llamadas, enviar mensajes, tomar fotografías, ver contenido multimedia y navegar por internet, todo, en un reducido tamaño. No obstante, el crecimiento de la industria tecnológica con la tendencia de dotar a todo como *Smart* (*Smartphones*, *Smartwatches*, *Smart TV*, *Smarthouses*, *Smartcities*), el desarrollo de software mediante la creación de aplicaciones y plataformas digitales, además de la digitalización social con la popularización de las redes sociales, ha traído consigo la formación de un océano digital que se encuentra en expansión constante. Dicho océano de datos conocido más comúnmente como Big Data se encuentra conformado por metadatos generados por toda la actividad que realizan los usuarios a través de la red.

Si bien el Big Data ha tenido presencia desde la década de los setenta con la construcción de las primeras bases de datos integradas en SQL (Structured Query Language)<sup>1</sup>, no fue hasta el año 2005 con la presencia de las redes sociales, la inclusión de lenguajes alternativos como NoSQL<sup>2</sup> y la web 2.0 cuando el término comenzó a difundirse con mayor fuerza para tratar de describirlo. Por lo que, a pesar de que el término Big Data se ha vuelto bastante popular en la era digital, aun hoy en día encontrar una definición precisa que logre explicarlo puede ser complicado, ya que la mayoría de las aproximaciones que se han hecho para describir, qué es Big Data, lo colocan con cuatro características en común; volumen, velocidad, variedad y veracidad, haciendo énfasis en la primera característica, lo que ha provocado el escepticismo de algunos expertos en torno al Big Data.

---

<sup>1</sup> El Structured Query Language (SQL por sus siglas en inglés) es un lenguaje utilizado en programación, diseñado para administrar y recuperar información de sistemas de gestión de bases de datos relacionales.

<sup>2</sup> NoSQL o también conocido como No solo SQL, en un conjunto de sistemas de gestión de bases de datos que difieren del modelo de gestión de datos ya establecido de SQL, refiriéndonos al Sistema de Gestión de bases de Datos Relacionales.

De tal manera que, el presente trabajo de investigación tiene como objetivo central determinar la utilidad que tiene el Big Data para la sociología, pues la migración de los medios de comunicación tradicionales hacia plataformas digitales y la popularización de las redes sociales, ha derivado en la apertura de nuevos campos de estudio que pueden ser explotados por la sociología, debido a que hoy en día la normalización de actividades como la publicación de fotos o videos en redes sociales, compartir ubicaciones e inclusive valorar la calidad de los servicios ha permitido que la sociedad se sienta más inmersa en la era digital desarrollando identidades virtuales construidas a partir de la interacción de los individuos a través de la red, y para ello, es necesario identificar las ventajas que tendría el análisis de Big Data como recurso de investigación en la sociología, considerando las técnicas de búsqueda, recolección y análisis de información; asimismo, conocer las implicaciones (desventajas/obstáculos) que tiene para un sociólogo realizar un análisis de datos a través de Big Data.

Ahora bien, tomando en cuenta que los grandes datos se encuentran conformados por conjuntos de conjuntos de datos que pueden estar estructurados, semiestructurados o no estructurados dependiendo del tipo de dato<sup>3</sup> que esté relacionado a esos conjuntos, la pregunta recae en ¿cuál es la utilidad que tiene Big Data en la sociología? Por lo que, si bien la respuesta yace en la calidad y la cantidad de información que es posible aprovechar analizándolo, actualmente mencionar Big Data en una época en donde la mayor parte de las cosas se rigen por tendencias, pueden ser encontradas por internet, o bien, se encuentran conectadas a él. Resulta curioso entender como el internet y toda la información que circula en ese océano digital se han vuelto una de las herramientas más indispensables de la sociedad actual, a tal grado que la industria tecnológica e informática no han sido capaces de gestionar el gran volumen de datos generados

---

<sup>3</sup> 1. Las búsquedas web y las redes sociales; 2. La información enviada entre dispositivos por el protocolo de *Machine to Machine* (M2M) o *Machine Learning*; 3. Los datos sobre transacciones bancarias o de la bolsa de valores (sistema financiero o mercado bursátil); 4. Datos biométricos (huella dactilar, escaneo de retina); y por último 5. Los datos generados por humanos como documentos electrónicos, así como los registros vocales de notas voz o llamadas.

diariamente, y aunque la tendencia apunta a que el volumen de datos seguirá incrementándose en vez de disminuir, esos datos “pueden ser de gran importancia principalmente en la toma de decisiones” (Camargo-Vega et al. 2015: 65), de tal forma que el Big Data cobra importancia no tan solo para las empresas que vieron un alto potencial en la explotación de los datos, sino también para el sector académico que intenta descifrar las iteraciones del comportamiento social a través de los conjuntos de datos, ya que se ha vuelto una respuesta oportuna para la resolución de problemas. Es decir, los grandes datos pueden ser utilizados esencialmente como un recurso de análisis para medir el comportamiento social como se ha demostrado en diversas ocasiones con el manejo de las campañas electorales (refiriéndonos a Cambridge Analytica), publicidad, índices de retención de clientes y el funcionamiento de algoritmos de recomendación en plataformas digitales (refiriéndonos a Amazon).

Como se mencionó anteriormente, uno de los ejemplos más recientes sobre información extraída del *Big Data* es el caso *Cambridge Analytica y Facebook*, en el que se dio a conocer que con ayuda de un *quiz*<sup>4</sup> en la red social pudieron obtener algo más de 87 millones de datos de usuarios, resultando en una de las filtraciones más grandes de la historia de internet, pero que permitió a la consultoría analizar información y crear públicos objetivos mediante análisis de Big Data con información obtenida de redes sociales, para después bombardearlos con anuncios, noticias falsas, cifras maquilladas o cualquier información que pudiese alterar su toma de decisiones como ocurrió con la tendencia de voto en el proceso electoral de Estados Unidos en 2016. No obstante, aunque es posible considerar los casos como el de *wikileaks* con la filtración de documentos de índole gubernamental o los *Panama Papers* con la exposición de los paraísos fiscales, como una prueba del alcance que puede tener la utilización del Big Data, es más del resultado de organizaciones *hacktivistas* que pretendían hacer visibles los actos de corrupción o crímenes de gobiernos involucrados. Lo que sí resulta directamente vinculado con el Big Data es

---

<sup>4</sup> Un Quiz es un cuestionario o formulario con un formato similar al de un juego de destreza o habilidad mental, en el que los jugadores responden a preguntas correctamente para tener la puntuación más alta o bien para conocer aspectos sobre su personalidad. A diferencia de un test, estas pruebas suelen ser más rápidas por lo que contienen pocas preguntas.

el periodismo que se derivó de la aparición de esos documentos denominado ahora *periodismo de datos*.

Por otro lado, este no es el único caso en el que se ha recurrido al Big Data, siendo el marketing y la publicidad otros sitios en los que se destaca su utilización, debido a la versatilidad de sus aplicaciones, que a diferencia de *Cambridge Analytica* en donde evaluaron las posibilidades de cambiar la toma de decisiones de la población, con el marketing se evalúa la retención y la pérdida de clientes a través de modelos predictivos con la intención de encontrar o bien identificar un sector de interés para las grandes marcas, sumando también la gestión empresarial que permite evaluar el riesgo de inversión a través de algoritmos diseñados para estimar las ganancias y las pérdidas que pueda tener una empresa; asimismo, los bancos también suelen usar este análisis, utilizando la geolocalización del cuentahabiente, la fecha y hora en que es realizada una transacción, además de la plataforma en que realizó para prevenir las actividades comerciales fraudulentas la mayoría de las ocasiones.

Hasta este punto, todo indicaría que los grandes datos podrían ser uno de los recursos más valiosos que tendría el investigador social en la época actual, ya que "...esta explosión de "grandes datos" está transformando la manera en que se conduce una investigación adquiriendo habilidades en el uso de Big Data para resolver problemas complejos relacionados con el descubrimiento científico, investigación ambiental y biomédica, educación, salud, seguridad nacional, entre otros" (IBM, 2012, 8). De tal forma que puede proveer de información específica al usuario con datos que no se obtenían fácilmente, como las preferencias electorales en tiempo real, la efectividad de una política pública, la eficacia de una campaña publicitaria a través de la *social media*, documentos académicos o aportes científicos de universidades en otros países, los censos de población, índices socioeconómicos de manera regional, así como consultar en tiempo real las decisiones tomadas por el congreso sobre los temas de la agenda pública, entre otros.

En este sentido, el uso Big Data por sus características, ha permitido a los investigadores revelar diversos aspectos como datos o variables dentro de su investigación que tal vez les habría tomado años descubrir sin el uso de estas herramientas, haciendo posible que el propio investigador pueda sopesar nuevas perspectivas basadas en los resultados obtenidos de su análisis de tal forma que el investigador pueda encontrar el valor real al que esta tratando de llegar. Lo que resulta enriquecedor para la sociología, ya que de esta forma no se encontraría limitando su campo de estudio a variables tradicionales, sino que complementa sus ejes de análisis e iría construyendo nuevas categorías, debido a que los resultados obtenidos por Big Data dependen más de la calidad de la información que pueda ser recolectada y no de la cantidad de información que pueda ser procesada.

Y aunque, se puede ir vislumbrando la utilidad que tiene el Big Data en la sociología, aún resta por conocer ¿cuáles son las implicaciones que tendría la incorporación de los análisis de Big Data a la sociología? Que obviando las características generales que determinan el funcionamiento de Big Data, tal vez el obstáculo más notorio sería comprender cómo se podría realizar un análisis de Big Data con una perspectiva sociológica, pues esto puede hacer suponer que el investigador debe contar con un vasto conocimiento en informática, y aunque esto puede ser un factor que influya para que pueda realizar el análisis, lo cierto es que no es determinante.

Por otro lado, la razón por la cual se eligió el Big Data entre tantos objetos de estudio posibles que tuvieran una mayor afinidad con la sociología, se debió principalmente a la inquietud por encontrar una forma en que la sociología pueda analizar los datos que son generados diariamente en el ciberespacio (redes sociales, sitios web o aplicaciones móviles) para usarlos como una herramienta de investigación complementaria a las ya existentes en la sociología; asimismo, al ser un tema que actualmente se encuentra modificando los paradigmas sobre cómo se analiza la información, es necesario conocer cómo aprovechar toda la infraestructura del Big Data de manera que actúe como catalizador para obtener nuevas perspectivas sobre fenómenos sociales preexistentes e inclusive lograr observar la evolución de

fenómenos sociales más recientes como la digitalización social o las identidades virtuales, sólo por mencionar algunos ejemplos.

Cabe aclarar que para el desarrollo de este trabajo debido a los pocos trabajos que se tienen sobre el Big Data aplicados desde la perspectiva sociológica o más general a las ciencias sociales, fue necesario consultar fuentes directas que permitiesen explicar Big Data, por lo que al ser fuentes especializadas la mayoría de los documentos que se utilizaron como apoyo para fundamentar el apartado teórico de la tesis fueron reinterpretados de tal forma que pudiesen ser entendibles para el lector sobre la forma que puede ser aplicado el Big Data en la sociología. Estos documentos incluyen artículos, glosarios, boletines o documentos científicos de empresas especializadas en el *cloud computing*<sup>5</sup>, sistemas financieros, mercadotecnia e inclusive de organismos especializados en el sector salud. En otras palabras, se encuentra sustentado en un análisis de tipo documental.

Aunado a lo anterior, el presente trabajo puede ser considerado como una investigación de carácter explicativa-exploratoria. Es decir, se considera exploratoria debido a la poca información que hay de Big Data relacionada con la sociología, asimismo, se considera explicativa debido a que la investigación tiene la intención de develar o encontrar un vínculo favorable del Big Data que pudiera ser explotado desde la sociología.

En este sentido, la tesis ésta compuesta por una introducción, tres capítulos en los que se desarrolla la utilización del Big Data como parte de la sociología y, por último, un apartado de conclusiones en el que se detallan las ventajas y las desventajas que puede tener la aplicación de los grandes datos como parte de los quehaceres sociológicos. Es así que, en la parte introductoria se engloban los motivos por los cuales se decidió estudiar al Big Data sobre los otros objetos de estudios que pudieran tener una mayor afinidad con la sociología, asimismo en esta parte, se

---

<sup>5</sup> Gartner inc. define *cloud computing* o *computo en la nube* como un “...estilo de computación en el que las capacidades de TI escalables y elásticas se brindan como un servicio utilizando tecnología de internet”, es decir, se entenderá *cloud computing* como un servicio tecnológico que permite el acceso remoto a almacenamiento de archivos, software o procesamiento de datos por medio de internet a través de una computadora personal o un servidor local. Para más información puede consultar <https://www.oracle.com/mx/cloud/what-is-cloud-computing/#link1>

plantean los objetivos y las preguntas que fungirán como guía a lo largo de la investigación además de la metodología empleada.

El capítulo uno tiene la intención de explicar qué es Big Data por lo que se describe el concepto, tomando como punto de partida el almacenamiento digital y los sistemas de información para establecer los antecedentes que fungieron como un catalizador para el surgimiento de los grandes datos, asimismo, se presentan las definiciones más utilizadas de Big Data con el propósito de esclarecer el término y construir una definición que pueda ser entendible desde la sociología, adicionalmente a esto, se establecen las características que lo componen, los tipos de información que se pueden encontrar (que adelantando un poco pueden ser datos estructurados, semiestructurados y no estructurados subdivididos en categorías)

Por su parte, el segundo capítulo se enfoca en determinar la importancia de analizar el Big Data desde una perspectiva sociológica, por lo que toma como referencia algunos de los casos en los que Big Data ha sido aplicado en el mundo real presentados desde diversas perspectivas como política, publicidad, medicina o informática como una forma de destacar los posibles avances que pueden llegar a tener si se elige adoptar el Big Data como parte de la sociología, en donde además se presentan los niveles de profundidad existentes en Big Data, los cuales no se deben confundir con los tipos de información que se pueden encontrar, aunque están relacionados. Por último, en este segundo capítulo también se establecen tres aspectos esenciales (la información que se desea obtener, el propósito de la investigación y el beneficio de los datos obtenidos por un análisis de Big Data, referente a la calidad de los datos) que como investigadores debemos conocer para determinar la importancia de los grandes datos en la sociología.

Si bien el segundo capítulo tiene la finalidad de mostrar la importancia de analizar Big Data, el tercer capítulo se encuentra enfocado en determinar la utilidad que tiene el Big Data para con la sociología, por lo que se establecen varios de los elementos que permitirán al investigador/sociólogo poder decidir si la utilización del Big Data puede ser viable dentro en una investigación de carácter sociológico. De tal manera

que, en este capítulo el lector se podrá encontrar con elementos como la construcción de ecosistemas digitales que funcionan como un medio para recolectar datos de formas no convencionales para la sociología, por ejemplo, a través de medios digitales, así como una mejor forma de administrar los flujos de datos que en ocasiones puede ser complicado si no se está acostumbrado a ellos. Asimismo, se explica cómo la datificación (aumento masivo de datos) se ha vuelto un punto crucial dentro de la sociología para estudiar nuevos paradigmas entre los que se encuentran las identidades virtuales como una nueva representación social.

Es decir, dentro de este tercer capítulo se muestran los aportes que traería consigo adoptar el Big Data como parte de la sociología, tomando en cuenta que la utilización de grandes datos dejaría de limitar al investigador a seguir las corrientes tradicionales. Adicionalmente a eso, se describen las similitudes que los distintos tipos de analítica presentes en Big Data con el proceso de investigación seguido en la sociología. Relacionado con esto, también se describen los tipos de análisis que pueden ser realizados con Big Data y las similitudes que presentan con los procesos de análisis empleados en la sociología.

Por último, como apartado de conclusiones se presentan las ventajas y desventajas que trae consigo incorporar el Big Data dentro de la sociología, adicionalmente se establecen algunas recomendaciones para contrarrestar algunas de las desventajas.

# **Capítulo 1. El almacenamiento digital y los sistemas de información como origen del Big Data**

Antes de comenzar con la definición de *Big Data* proporcionada por académicos, empresas u otras organizaciones es necesario entender la composición de este. Si bien, el Big Data es un conjunto de conjuntos de datos, adelantándome un poco con las definiciones tradicionales, existen dos rasgos peculiares que han permitido que estos datos masivos puedan ser considerados como tal, junto a sus características particulares. Estos son: los sistemas de información y las unidades de almacenamiento que como en una computadora, PC u ordenador son segmentados en el hardware y software. Pero, en este caso aplicado al Big Data, la combinación de estos dos elementos da como resultado todo el entramado de grandes volúmenes de información que lo compone.

Por lo que este primer capítulo tiene la intención de explicar el concepto Big Data tomando como punto de partida los sistemas de información y el almacenamiento digital, debido a que estos dos elementos son el preámbulo para entender el surgimiento de los grandes datos, ya que, sin ellos, actualmente nos sería imposible llevar a cabo un análisis de cientos de millones de datos esparcidos en el ciberespacio. Asimismo, en este primer capítulo se presentan las definiciones más utilizadas de Big Data con el propósito de asimilar el término y definirlo desde la sociología, adicionalmente, se establecen sus características y los tipos de datos que se pueden encontrar, que pueden ser estructurados, semiestructurados y no estructurados.

## **1.1 El almacenamiento digital y los sistemas de información**

Desde el inicio de la historia, los seres humanos han tenido la necesidad de compartir y almacenar experiencias adquiridas de su entorno para que futuras generaciones puedan hacer uso de esos conocimientos, por ello, es que a través del tiempo han surgido diversos formatos, que han hecho esto posible. Siendo en un principio formas rudimentarias que estaban destinadas a registrar los

acontecimientos más relevantes. Pero que, con el paso de los años fueron sofisticándose, adquiriendo una estructura de modo que se pudiese consultar la información de manera eficaz.

Más tarde, la carrera aeroespacial marco la pauta para el desarrollo de las primeras computadoras, que suscitaron una transformación en la forma de almacenar información, debido a que la memoria requerida, además de la potencia de procesamiento, era fundamental para cumplir con su cometido de colocar al hombre en la luna y registrar con éxito ese acontecimiento. Dichas unidades de almacenamiento fueron evolucionando conforme a las necesidades de los usuarios en términos de capacidad de almacenamiento, poder de procesamiento y velocidad de escritura, lo que orilló a los desarrolladores a diseñar recipientes que pudieran acumular una mayor cantidad de información en un menor espacio para ser transportado, lo que derivó en un principio con los llamados disquetes entre 1971-1999 que pudieron almacenar poco más de 1.5 MB que para la época fue sumamente de utilidad. Pero, poco a poco fue remplazado con los CD's que tenían 900MB y posteriormente con los DVD's con poco menos de 9GB. Finalmente, los pendrives o memorias USB terminaron por hacer una revolución en cuanto al almacenamiento portátil, debido a que hoy en día es posible almacenar hasta un Terabyte de información en dispositivos que caben en la palma de la mano (véase Figura 1.)

Por otra parte, el almacenamiento portátil no fue el único que tuvo una transformación, la demanda de espacio dentro de los ordenadores también efectuaron un cambio importante en la *ROM*<sup>6</sup>, ya que los discos duros mecánicos (HDD) en sus presentaciones de 3.5 y 2.5 pulgadas pasaron a ser a estado sólido (SSD) con nuevos protocolos que aumentan la velocidad de escritura. Al mismo tiempo fue necesario un aumento en la memoria *RAM*<sup>7</sup> de los dispositivos para

---

<sup>6</sup> Es el sistema de almacenamiento interno de los dispositivos llámese ordenador, teléfono, tableta o cualquier otro dispositivo que pueda guardar información propia.

<sup>7</sup> Es el almacenamiento temporal que tienen los dispositivos, con ella se guarda información mientras esta en uso y posteriormente se pasa al disco duro (ROM) o se elimina. Este tipo de memoria acelera la apertura de programas o archivos.

compensar la capacidad de procesamiento. De esta forma se aumentaba la cantidad de ciclos por segundo en escritura y visualización de información.

La necesidad de acopiar inmensas cantidades de información ha resultado en la creación de bancos de datos donde es acumulada, para después ser ordenada y analizada. Sin embargo, estos conjuntos de datos que podían ser transportados por disquetes, discos, pendrives o simplemente guardado en un ordenador, se vio rebasado hasta el punto de crear un espacio a gran escala e intangible para el usuario final, esto debido a toda la información generada diariamente a través de las búsquedas en la red, minutos de video subidos a YouTube, imágenes publicadas en Facebook, estados escritos en Twitter, o bien, las ubicaciones compartidas por cualquier servicio. Al final de cuentas, con el tiempo la expansión de memoria junto con el aumento en la capacidad de procesamiento no fue suficiente para los bancos de información dentro de los ordenadores convencionales, dejando en claro que el almacenamiento de las computadoras era muy limitado. Optando así, por el desarrollo de sistemas de almacenamiento remotos con mayor capacidad.

**Figura 1. La transformación de las unidades de almacenamiento.**



Fuente: Retomada de unocero.com

La llegada de los sistemas de información como algo complementario para buscar, categorizar, ordenar y analizar información fue el siguiente paso en la organización y la construcción de bases de datos. Pero, exactamente ¿qué es un sistema de información? En un primer acercamiento para entender un sistema de información (SI) es que funciona como un conjunto de componentes que tienen la tarea de recolectar, procesar y almacenar información en la que se puedan apoyar las empresas u organizaciones para tener un mejor control sobre la información que ingresan y así tener una mejor toma de decisiones respecto a los datos.

“Un sistema de información es considerado como sistema, automático o manual, para agrupar, procesar, transmitir y diseminar datos que representan información para el usuario” (Gonzalez-Longatt,2007:1). Retomando la definición en administración, este mismo sistema es utilizado comúnmente en bibliotecas para la consulta de libros y que se encuentra presente mediante ficheros, en donde se le asigna un código o clave dentro de un catálogo que resguardará esa información. Sin embargo, no es la única definición existente sobre los sistemas de información, puesto que desde la perspectiva de las telecomunicaciones también podría ser considerado como “un sistema interconectado o un subsistema de equipos de computación o telecomunicaciones que es usado en la adquisición, almacenamiento, control, presentación, comunicación, intercambio, transmisión o recepción de voz y/o datos, e incluye *software*, *firmware*, y *hardware*” (Gonzalez-Longatt, 2007: 1). En cualquiera de los dos aspectos, el diseño de un sistema que buscase y mostrase información exacta dentro de un almacenamiento de gran tamaño, permitió que el manejo de la información fuese mucho más práctico, haciendo posible que se pudiese procesar, catalogar y analizar información acumulada en ese océano extenso de metadatos sin estructura que es internet. Aplicado a la sociología estos sistemas de información más allá de ser subconjuntos de sistemas interconectados y en comunicación, se refiere a la posibilidad del investigador para tener un mejor control de los datos dentro del universo muestral. De esta manera evita, en la medida de lo posible, sesgos de información que pudieran surgir al momento de minar o recolectar datos, aumentando el valor de la información para su investigación.

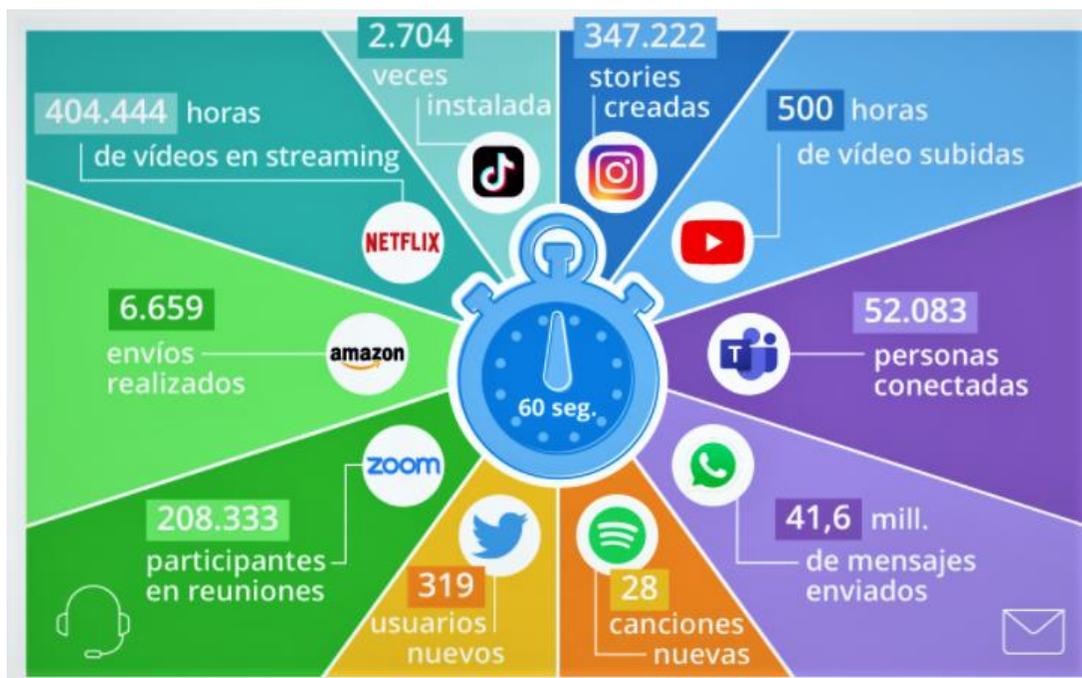
Es posible que, a estas alturas, el lector ya se haya percatado que el Big Data se encuentra diseñado con base en un sistema de información complementado a un espacio de almacenamiento físico, muy similar a internet con su sistema de ficheros y nodos de información que se ejecutan al mismo tiempo y envían información a las pantallas de los ordenadores. Pero ¿cómo funcionan los sistemas de información? Digamos que la manera más sencilla de explicar el funcionamiento de un sistema de información es como lo señalaba Durkheim con la división del trabajo social o con las líneas de producción propuestas por Henry Ford. Es decir, el modelo de producción Fordista integra estaciones de trabajo que se encargan de ensamblar una parte del vehículo; cada una de estas estaciones conoce el trabajo que debe realizar (ensamblar el chasis, colocar el motor, soldar la carrocería, colocar la electrónica, vestiduras, asientos, etc.). Sin embargo, ninguna de las estaciones puede realizar el trabajo del otro, debido a que no está capacitado para hacerlo como lo señala el esquema Durkheimniano de la división del trabajo social.

En el caso de los ficheros que se ejecutan en internet sucede algo parecido, cada fichero, cada nodo contiene la información que posee el otro. Por lo que, a diferencia de los ejemplos antes mencionados, estos ficheros son una copia fiel del original, de esta manera pueden acceder a una parte de la información de forma más rápida. Es decir, muestra la información desde el servidor más cercano del que provino la solicitud y, de esta manera es posible tener un mayor control con el tráfico de datos. En otras palabras, el desarrollo de un sistema de información al igual que con el tráfico de bytes en internet ayuda a tener un mejor y mayor control sobre la información que puede ser guardada.

Las empresas u organizaciones suelen utilizar estos mismos sistemas de información para monitorear su tráfico de datos. Por ejemplo, en el caso de Google una empresa dedicada a la tecnología, que además es pionero en el cloud computing y desarrollador de software multiplataforma, se mantiene generando cantidades exorbitantes de información, basta con decir que son realizadas aproximadamente 3.8 millones de búsquedas por minuto. (véase figura 2.) Es más que obvio que tal cantidad de información no puede ser almacenada dentro de una

computadora como la que la mayoría de las personas tienen en sus hogares, obligando a los punteros de tecnología a la creación de espacios como los Data center cuya infraestructura está compuesta por un espacio similar a una bodega, edificio, bunker donde resguardan las unidades de almacenamiento complementado a un sistema de información integrado a los ordenadores ensamblados dentro.

**Fig.2 Información en internet cada minuto**



Fuente: Retomada de Statista con actualización al primer semestre del 2020, disponible en <https://es.statista.com/grafico/17539/datos-creados-online-en-un-minuto/>

Hoy en día hablar sobre el Big Data también supondría hablar de la infraestructura que hay detrás de este; como los *Data center*, el *Data warehouse*, el *Hosting*. Básicamente edificios ubicados en diferentes partes del mundo o servicios de cómputo virtual que almacenan toda la información generada en la World Wide Web. Pero, el desarrollo de aplicaciones, la integración de las redes sociales como algo de la vida cotidiana, la popularización de internet, el aumento constante de información generada digitalmente, la falta de espacio disponible en los dispositivos

y la velocidad en que son escritos los datos, ha llevado a que la sociedad recurra a la integración de servicios especializados en almacenamiento remoto como los de Google, Microsoft, Apple, Oracle o IBM solo por mencionar algunos, haciendo que el almacenamiento ya no sea expresado en Megabyte, Gigabyte o Terabyte y cambie a unidades como Petabyte, Exabyte o Zettabyte.<sup>8</sup> sumando toda la información que es guardada diariamente.

## **1.2 Los Data warehouse y Data lake; el cauce de información para Big Data**

Hasta este punto el lector ha podido ver un poco el funcionamiento que tiene el Big Data en cuanto a su infraestructura; con la integración de las unidades de almacenamiento y los sistemas de información como son los Data center. Sin embargo, a pesar de que la integración de los centros de datos ha permitido que se pudiese aumentar la capacidad de almacenamiento sin preocuparse por el espacio dentro de los dispositivos. La categorización, el procesamiento y la agrupación de datos juegan un papel importante dentro de estas construcciones. Si bien, el lector ya intuye que la información que se produce en internet es guardada dentro de los centros de datos, este apartado tiene como fin explicar brevemente la forma en que el *Data warehouse* y el Data Lake funcionan como un río que alimenta de información al Big Data. Cabe aclarar que estos dos conceptos son formas en las que se puede recopilar información, por lo que la manera en que trabaja difiere principalmente en su arquitectura de diseño.

Si bien una de las particularidades que tiene el Big Data es que puede explotar casi cualquier fuente de información, el lugar donde se alimenta este monstruo informático es precisamente de los Data warehouse y los Data Lake, que resultan ser las fuentes de las que se extrae más información para los análisis de Big Data. Sin embargo, esto no exime a los analistas o al sociólogo construir sus propios bancos de datos con características propias mediante diversas técnicas de

---

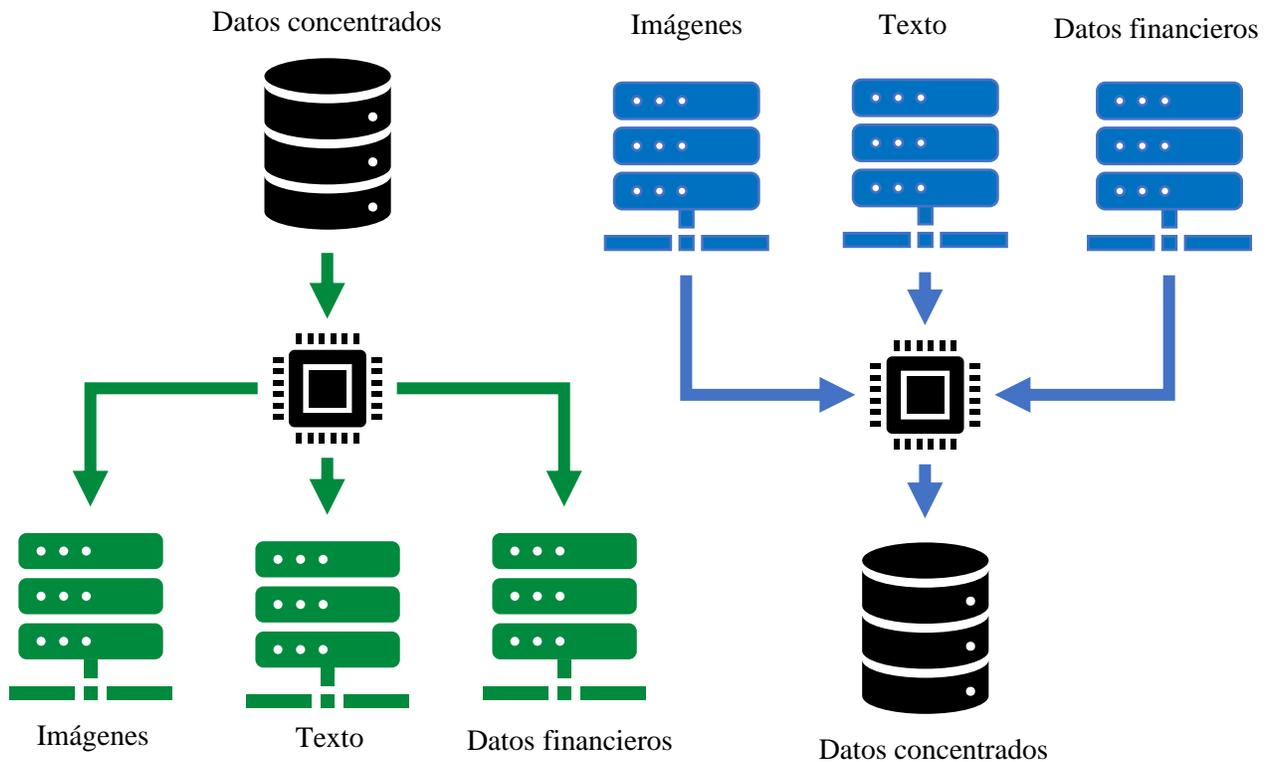
<sup>8</sup> Un Zettabyte es el equivalente de un sextillón de bytes (representado por la escala numérica corta) y mil trillones bytes (en la escala numérica larga), o lo que es igual a  $10^{21}$  representado por notación científica.

recolección de datos, como la minería o el farreo de datos. Pero ¿Qué es un *Data warehouse*? ¿qué es un *Data Lake*? y ¿por qué son un cauce de información para el Big Data? bien, un *Data warehouse* "... es un repositorio unificado para todos los datos que recogen los diversos sistemas de una empresa. El repositorio puede ser físico o lógico y hace hincapié en la captura de datos de diversas fuentes sobre todo para fines analíticos y de acceso" (powerdata, 2014). Es decir, imaginemos un edificio; dentro de este edificio hay cientos de computadoras acomodadas en torres con gabinetes funcionando 24/7 durante los 365 días del año; cada ordenador se encuentra en una ingesta constante de información que es recolectada y agrupada por las empresas sobre distintos aspectos como la información de los usuarios, la calidad de los servicios y las transacciones que realizan solo por mencionar algunos de ellos. Por otra parte, un *Data Lake* "... es un repositorio de almacenamiento que contienen una gran cantidad de datos en bruto y que se mantienen allí hasta que sea necesario. A diferencia de un *Data warehouse* jerárquico que almacena datos en ficheros o carpetas, un *Data Lake* utiliza una arquitectura plana para almacenar los datos" (powerdata, 2014). En otras palabras, los *Data warehouse* se encuentran compuestos principalmente por datos estructurados o semiestructurados como lo pueden ser aspectos específicos determinados por un algoritmo de *machine learning*, haciendo que el cauce de información que puede ser obtenida de esa fuente deba ser filtrada para determinar el valor de los datos. Caso contrario con el *Data Lake* que los datos son almacenados sin tratamiento alguno o categorización por lo que puede suponer bases de datos mucho más extensas, con mayor dificultad, pero, con una mayor riqueza en cuanto a su contenido. (véase figura 3.)

Decir que el *Data warehouse* y el *Data Lake* son un cauce de información para el Big Data es precisamente porque la forma en que una palabra, una oración, o una frase ya tabulada dentro de la base de metadatos puede convertirse en una potencial fuente de información. Nuevamente pondremos de ejemplo a Google, esta compañía recolecta información de los usuarios, las búsquedas que realizan, su geolocalización, los dispositivos que utiliza, la música que escucha, los videos que reproduce, los sitios web que visita, etc. Por lo que algunas personas confundirían estas acciones con espionaje y aunque roza una línea sumamente delgada, lo cierto

es, que Google realiza un seguimiento de sus usuarios para ofrecer de manera focalizada e incluso personalizada sus servicios. De la misma forma que ocurre con otras compañías como Apple, Amazon o Microsoft sólo por mencionar algunas.

**Figura 3. distinción gráfica de un Data warehouse y un Data Lake**



Fuente: Elaboración propia con base en las definiciones de powerdata e IBM

Nota: En la izquierda se puede apreciar el data warehouse y de lado derecho el Data Lake\*

Resulta curioso entender que el internet se haya vuelto una de las herramientas más indispensables de la sociedad actual y con ello, toda la información que circula en él se haya vuelto el objetivo de empresas u organizaciones, concibiendo la apertura de nuevos campos de acción relacionados a esos impresionantes volúmenes de datos. No obstante, no se debe perder de vista que empresas como Apple, Amazon, Google, Facebook, Twitter, IBM, Oracle o cualquier otra que se añada en el sector tecnológico, tiene a su merced estos espacios confinados para

almacenar información y por ello, buscarán la forma de analizar esa fuente potencial de datos nombrados Big Data.

**Figura 4. Data center de Google en Eemshaven, Países Bajos**



Fuente. Retomado de Google, conoce nuestros centros de datos en <https://www.google.com/about/datacenters/>

### **1.3 ¿Qué es el Big Data?**

Si bien ya pudimos observar un poco la composición del Big Data en cuanto a su infraestructura y como se va integrando información para ser analizada, ahora toca conocer ¿qué es el Big Data? mediante las definiciones que ha dado la academia y el sector empresarial, mientras tanto, es seguro que lo primero que imagina el lector cuando se habla de este tema es una cantidad masiva de datos. Pero, lo cierto es que la manera más sencilla de interpretar el *Big Data* es reconocerla como una base de datos gigantesca que es actualizada diariamente y que se encuentra en expansión. Es decir, algo similar a una presa que acumula cientos de litros de agua y que es liberada cada cierto tiempo para evitar que colapse. Bueno, lo mismo ocurre con esta enormidad de base de datos, todos los días es llenada de información y con cada análisis realizado es vaciada un poco.

El constante cambio en la sociedad ha permitido que en las últimas dos décadas las personas tengan un mayor acercamiento al desarrollo tecnológico y una mayor inmersión digital, trayendo consigo la llegada de nuevas tendencias para dotar a todo con el concepto de inteligente o *Smart* (*Smartphones, Smartwatches, Smart TV, Smartcities*) con la pretensión de facilitar el estilo de vida de las personas. Todo esto con ayuda de un océano compuesto por una cantidad exorbitante de datos y al que los usuarios contribuyen a su crecimiento mediante las búsquedas en Google, la publicación de imágenes en las redes sociales o al compartir su ubicación en plataformas digitales. Y aunque el término Big Data es relativamente nuevo, la acción en sí de recopilar y almacenar grandes cantidades de información para su posterior análisis se ha realizado desde hace muchos años.

La generación de grandes cantidades de información ha dejado entrever que “hoy las compañías no saben qué hacer con el gran volumen de datos e información almacenada en diferentes medios o bases de datos, los cuales pueden ser de gran importancia, principalmente en la toma de decisiones” (Camargo-Vega et al. 2015: 65). Como suele ocurrir actualmente en el manejo de las campañas electorales, marketing político y/o empresarial, publicidad, o bien, la retención de clientes a través del análisis de Big Data en el caso de las empresas que ofertan un servicio. “Si se indaga el volumen de libros con el vocablo Big Data en inglés con la aplicación Ngram Viewer de Google, su pico estadístico se localiza en el año 2000, lo que contribuye a darnos una idea de que su generalización es relativamente reciente” (Meneses, 2018; 418). Los primeros indicios del concepto datan de los años 70’s con las primeras bases de datos en SQL y que posteriormente en el año 2005 con el boom de las redes sociales fue desarrollado Hadoop, uno de los primeros sistemas en open source para el público en general que tenía como función principal almacenar y procesar estos grandes volúmenes de datos, popularizando también el lenguaje en NoSQL, haciendo que el crecimiento del Big Data fuese acelerado debido a la factibilidad en cuanto al costo-beneficio que tenían los usuarios, aminorando el costo para almacenar grandes cantidades de información.

Con el avance de la informática y el desarrollo de *software* se generó la inclusión de nuevos conceptos como *Inteligencia Artificial (IA)* que permite a los usuarios usar comandos de voz para controlar sus dispositivos o inclusive hacer reservaciones en algún restaurante sin tener que llamar; así como el *Internet de las cosas (IoT)* relacionado con la domótica, un campo de desarrollo de la ingeniería con el que establece la asistencia y la interconexión de todos los dispositivos conectados a internet dentro del hogar para monitorear o realizar acciones preestablecidas por el usuario; y por último, la creación de algoritmos a través de *Machine Learning* para identificar principalmente patrones de comportamiento o la selectividad de información. Cabe resaltar que entre estos tres conceptos existe una simbiosis para su correcto funcionamiento, pero que deja claro que ya los datos no tan solo son generados por humanos sino también a través de los gadgets que usamos día con día.

Aunque suene pretencioso decirlo, la construcción de la infraestructura que hay detrás del almacenamiento masivo de información como los Data center y los Data warehouse son el parteaguas para el inicio del hosting y el cloud computing, esta última conocida comúnmente como la nube, abriendo las posibilidades para que el Big Data pueda desarrollarse. Puesto que, al ser un servicio de cómputo virtual que no depende totalmente de la computadora del usuario y puede ser escalable a medida que se requiera. Por lo que, con estos avances el Big Data puede ser aplicado para toda la información que no es procesada o analizada usando herramientas convencionales.

Algunas de las definiciones encontradas sobre este concepto establecen que “... son conjuntos de datos o combinaciones de conjuntos de datos cuyo tamaño (volumen), complejidad (variabilidad) y velocidad de crecimiento (velocidad) dificultan su captura, gestión, procesamiento o análisis mediante tecnologías y herramientas convencionales tales como bases de datos relacionales y estadísticas convencionales o paquetes de visualización, dentro del tiempo necesario para que sean útiles” (powerdata, 2014). Definición que resulta similar a la proporcionada por Oracle en donde menciona que “el Big Data está formado por conjuntos de datos

de mayor tamaño y más complejos, especialmente procedentes de nuevas fuentes de datos. Estos conjuntos de datos son tan voluminosos que el software de procesamiento de datos convencional sencillamente no puede gestionarlos. Sin embargo, estos volúmenes masivos de datos pueden utilizarse para abordar problemas empresariales que antes no hubiera sido posible solucionar” (Oracle, 2014).

En el caso de IBM “... se refiere más a la tendencia en el avance de la tecnología que ha abierto las puertas hacia un nuevo enfoque de entendimiento y toma de decisiones, la cual es utilizada para describir enormes cantidades de datos (estructurados, no estructurados y semi estructurados) que tomaría demasiado tiempo y sería muy costoso cargarlos a un base de datos relacional para su análisis” (IBM, 2012). Al igual que Gartner que considera el Big Data como “... activos de información de gran volumen, alta velocidad y/o gran variedad que demandan formas de procesamiento de información innovadoras y rentables que permiten una visión mejorada, toma de decisiones y automatización de procesos” (Gartner, 2012). Ejemplo de esto son las bases de datos hechas con sistemas de información como SQL y NoSQL, que además de la infraestructura de la que ya se ha mencionado, fueron el inicio del crecimiento del Big Data hasta su maduración con los procesos que son llevados a cabo actualmente en distintas disciplinas.

La mayoría de las aproximaciones que se han hecho para describir el término Big Data destacan que son una serie de conjuntos de datos con características particulares como; el volumen, la variabilidad o variedad y la velocidad principalmente, “... el término “big data” es una constante y, aun así, la mera noción sigue generando confusión. Esta palabra se ha utilizado para trasladar al público todo tipo de conceptos entre los que se incluyen grandes cantidades de datos, analítica de redes sociales, herramientas de última generación para gestionar los datos, datos en tiempo real y mucho más” (IBM,2012:1). Por lo que, a pesar de que se ha popularizado, hoy en día, aún es complicado encontrar una definición precisa que lo explique provocando cierto escepticismo, debido al énfasis que existe en el volumen de los datos que se almacenan para ser considerado como un Big Data.

Cabe resaltar que, “Big Data no va del todo dirigido al gran tamaño, sino que abarca tanto volumen como variedad de datos, velocidad de acceso y procesamiento” (Hernández Leal et. al, 2017: 4). Es decir, el Big Data como paradigma tecnológico ha transformado la forma de pensar y de tomar decisiones, inmiscuyendo el valor que ofrecen los datos en sí, adelantándose un poco a la importancia de analizar dichos volúmenes de información.

“Es importante entender que las bases de datos convencionales son una parte importante y relevante para una solución analítica. De hecho, se vuelve mucho más vital cuando se usa en conjunto con la plataforma de Big Data” (IBM, 2012: 2). Si bien es cierto que, el Big Data inspira un cambio en la forma de tomar decisiones y brindar soluciones dinámicas para los problemas actuales en la sociedad. También es cierto que dada la composición del Big Data como herramienta se cree que la mayoría de la información extraída de los almacenes de datos, o mejor dicho que alimenta Big Data, son procedentes de las redes sociales. Pero, al igual que se tenía escepticismo sobre la funcionalidad que tenían las Tecnologías de la Información y Comunicación (TIC) con el Big Data la “... evaluación teórica de este enorme caudal de información permite elaborar criterios teóricos generales para integrar los resultados obtenidos” (Forte, Pignuoli, Calise et al. 2012: 208). Visto desde la sociología, permitiría la creación de un ecosistema digital que contribuya en la forma de analizar los fenómenos sociales, la estructura social o el comportamiento, que en su momento fueron dilucidados por Weber, Durkheim o Bourdieu, sólo por mencionar algunos, pero que hoy en día son apreciados de forma distinta, es decir, han evolucionado. Por ello, el uso de técnicas no convencionales, como la minería o el farreo de datos, complementadas con las tradicionales como las encuestas, cuestionarios, entrevistas a profundidad, grupos focales, entre otras son una gran estrategia factible para poder analizarlos.

Sin duda alguna, al reconocer el Big Data como parte de esta era moderna en donde la mayor parte de las cosas se vuelven tendencia, pueden ser encontradas en internet, o bien, están conectadas a él. Se puede vislumbrar que: “La digitalización de prácticamente “todo” da lugar a nuevos tipos de grandes datos en tiempo real en

un amplio abanico de sectores” (IBM,2012: 3). Sumado a la tecnología y a las “... técnicas de análisis avanzado de hoy en día permiten a las empresas extraer conocimientos de los datos con un nivel de sofisticación” (IBM,2012: 3). Han jugado un papel sumamente importante dentro de la era digital, así como, también son piedra angular para el entendimiento del Big Data. Datos como los de streaming, geoespaciales o generados por sensores que divergen con las arquitecturas tradicionales de los Data warehouse y que por tal motivo no se encuentran normalizados.

#### **1.4 Las características del Big Data**

Más allá de su traducción literal al español como Datos Masivos, el Big Data posee características propias para que pueda ser considerado como tal, debido a que el concepto en sí es sumamente amplio y puede resultar ambiguo para quien habla de él. Si bien, en las definiciones que se tienen de Big Data se hace hincapié en las grandes cantidades de datos, no es lo único que lo caracteriza. Además del volumen de datos, el Big Data se encuentra definido por la Variedad o Variabilidad, la Velocidad y la Veracidad de los datos como principales componentes, aunque a dichos componentes se le pueden agregar otros que sean complementarios, dependiendo del investigador y/o empresa que esté realizando un análisis. Por lo que en este apartado tiene como finalidad esbozar las peculiaridades que posee el Big Data.

##### **1.4.1 Volumen**

Siendo la característica asociada con mayor frecuencia al entorno del Big Data; “el volumen hace referencia a las cantidades masivas de datos que las organizaciones intentan aprovechar para mejorar la toma de decisiones en toda la empresa” (IBM, 2012: 4). Con esta dimensión del Big Data se puede intuir que son extensas cantidades de datos desorganizados que se encuentran almacenados esperando ser procesados, categorizados y analizados. Pero que “registran un aumento significativo de información creada por personas y máquinas” (Camargo-Vega et.al, 2015: 66). Es decir, todo lo que se ha ido acumulando cada día desde diversas

fuentes de datos, además de la información proveniente de los Data center, Data warehouse, o los Data Lake, incorporando además información sobre transacciones comerciales, social media, información *machine to machine*, dan la posibilidad al Big Data de escalar su almacenamiento. De tal forma que, los datos siguen aumentando de manera exponencial, lo que establece que estos volúmenes varíen en función del sector, las necesidades del proyecto, la información recopilada e incluso de la ubicación geográfica.

Visto desde la perspectiva sociológica, saber utilizar grandes volúmenes de información significaría que es posible recolectar una mayor muestra correspondiente al estudio realizado con equipos de trabajo no tan grandes, al menos en un primer momento, debido a que la información es acumulada en un punto específico. Es decir, que la información se va centralizando en un almacenamiento local como un servidor *on premise*, o bien, uno remoto como el hosting o la nube.

#### **1.4.2 Variedad o variabilidad**

El volumen de los datos no es la única característica que tiene el Big Data, aunque es una de las que tiene mayor énfasis cuando se habla del tema. Sin embargo, la variedad o variabilidad también tiene un papel importante dentro del entorno de los grandes datos, debido a que “va muy de la mano con el volumen, pues de acuerdo con éste y con el desarrollo de la tecnología, existen muchas formas de representar los datos; es el caso de datos estructurados y no estructurados; estos últimos son los que se generan desde páginas web, archivos de búsquedas, redes sociales, foros, correos electrónicos o producto de sensores en diferentes actividades de las personas” (Camargo-Vega et.al, 2015: 66). Es decir, los datos son encontrados en formatos distintos y son extraídos de diversas fuentes.

Visto desde una perspectiva sociológica, la posibilidad de recabar diferentes clases de datos y englobarlos en un mismo volumen permite visualizar tendencias con mayor notoriedad. No obstante, los flujos de información circulando pueden responder a eventos coyunturales que incrementan las cargas de información de un

determinado fenómeno, aumentando a su vez la complejidad para analizar los datos, dificultando la manipulación de los bloques de información. Por ello, “la variedad tiene que ver con gestionar la complejidad de múltiples tipos de datos, incluidos los datos estructurados, semiestructurados y no estructurados” (IBM,2012: 4). En el caso del sociólogo se refiere más al control de su universo muestral cuando se enfrenta al fenómeno social estudiado, en donde al igual que las empresas debe “integrar y analizar datos de un complejo abanico de fuentes de información tanto tradicional como no tradicional” (IBM,2012: 4), dependiendo de los aspectos analíticos particulares del trabajo.

### **1.4.3 Velocidad**

Manipular grandes volúmenes de información proveniente de diversas fuentes no es sencillo si no se cuenta con las herramientas necesarias para procesar, catalogar y almacenar dicha información. Por lo que en este punto “Se refiere a la velocidad con que se crean los datos, que es la medida en que aumentan los productos de desarrollos de software (páginas web, archivos de búsquedas, redes sociales, foros, correos electrónicos, entre otros)” (Camargo-Vega et.al, 2015: 66). Es decir, la velocidad con la que son transmitidos los datos para catalogar la información a través de etiquetas o sensores.

“Hoy en día, los datos se generan de forma continua a una velocidad a la que a los sistemas tradicionales les resulta imposible captarlos, almacenarlos y analizarlos. Para los procesos en los que el tiempo resulta fundamental, tales como la detección de fraude en tiempo real o el marketing “instantáneo” multicanal, ciertos tipos de datos deben analizarse en tiempo real para que resulten útiles para el negocio” (IBM,2012: 4). Siendo así, que el poder de procesamiento sea un factor necesario para cumplir con el objetivo del Big Data, debido a que los datos siguen incrementando. Por lo que, el constante movimiento de los datos contribuye a la creación de nuevos datos casi en tiempo real. No obstante, esta característica se trata principalmente sobre la capacidad transferir grandes paquetes de datos con una menor latencia. Es decir, mejorar el tiempo de transferencia entre que se crea

la información, se almacena, hasta que se accede a ella, disminuyendo la correlación de pérdida/tiempo-información.

#### **1.4.4 Veracidad**

Considerar la incertidumbre de los datos es parte fundamental para la realización de un análisis de Big Data, debido a que brindaría una mayor confianza al investigador para presentar los resultados. “La veracidad hace referencia al nivel de fiabilidad asociado a ciertos tipos de datos. Esforzarse por conseguir unos datos de alta calidad es un requisito importante y un reto fundamental de big data, pero incluso los mejores métodos de limpieza de datos no pueden eliminar la imprevisibilidad inherente de algunos datos, como el tiempo, la economía o las futuras decisiones de compra de un cliente” (IBM, 2012: 5). En otras palabras, los datos de alta calidad no se refieren a la cantidad de datos analizados, sino a la información proporcionada por los mismos lo que determina su valor, es decir, los datos con una discrepancia mucho menor que el resto tienen un valor más alto.

No obstante, además de la autenticidad de la información recolectada, es oportuno considerar el aspecto legal, que, si bien la mayor parte de los Big Data son autenticados por datos que no comprometen la integridad de los usuarios, se debe tener en claro este aspecto. Desde la perspectiva sociológica tener un alto porcentaje de autenticidad en los datos permite tener resultados muchos más concretos que brinden una mayor confianza al investigador sobre su trabajo. Por supuesto esto se encuentra determinado principalmente por las fuentes de donde fueron recabados los datos, además de los instrumentos que se utilizaron para obtenerlos.

#### **1.4.5 Valor**

Aunque la calidad de los datos puede determinar su valor, no es lo único que lo hace. Algunas de las características sirven para considerar que tan valioso es un Big Data o un estudio realizado por Big Data, es fijado por la utilidad de la información, es decir, debe contestar a la pregunta ¿para qué sirven estos datos?; además de eso, otro de los aspectos que pueden establecer el valor intrínseco de

los datos es la procedencia. Por ejemplo, en el año 2012 durante la campaña electoral de Barack Obama fue realizado un minado de datos a través de redes sociales para determinar el sector de la población al que debía enfocarse para renovar su mandato presidencial, por lo que solicitó el consentimiento de la gente para acceder a sus datos sin infringir las normas de privacidad de los usuarios. Caso contrario el sucedido en 2016 con Cambridge Analytica donde a través de un *quiz* en una red social recabo información para fijar público objetivo y bombardearlo de publicidad con la intención de cambiar su voto.

#### **1.4.6 Otras dimensiones consideradas**

Además del volumen, la variedad, la velocidad, la veracidad y el valor de los datos, existen otras dimensiones del Big Data, que, a pesar de no estar consideradas como parte de las características generales son agregadas de manera particular por los investigadores, analistas o las empresas, dependiendo de los requerimientos de sus estudios. Un ejemplo de ello es la empresa y consultoría SAS que consideran la complejidad como una dimensión más para realizar un análisis de Big Data.

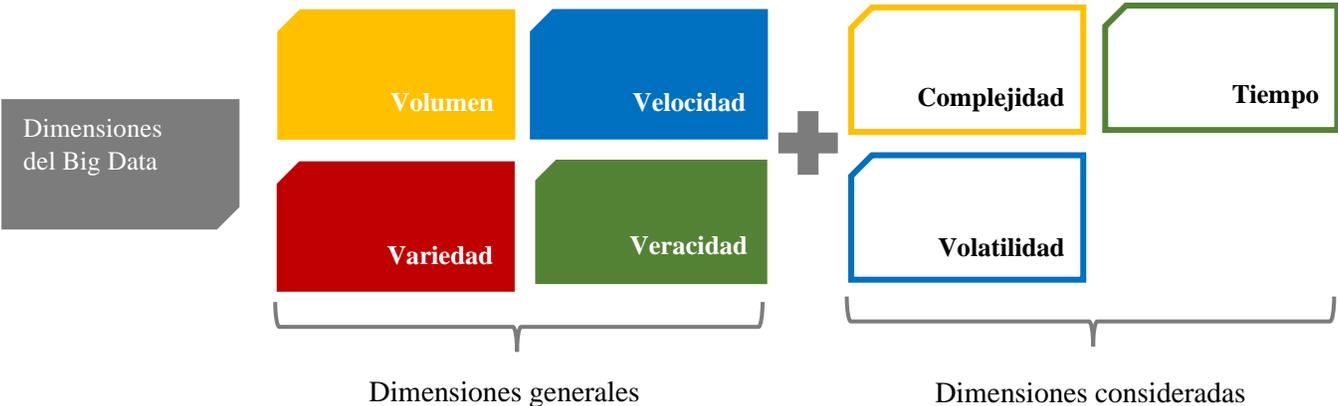
Se refieren a la complejidad de los datos, cuando los datos provienen de múltiples fuentes dificultando el procesamiento, es decir, los datos al ser recolectados no poseen en mismo formato, situación habitual en Big Data, ya que pueden contener un sin número de archivos diversos que es necesario organizarlos para que puedan ser analizados. De esta forma, es posible tener más claro el potencial de los datos y vislumbrar el entramado de la información para formar la mayor cantidad de correlaciones, jerarquías o vínculos, manteniendo bajo control la volatilidad de estos.

A punto personal, es posible identificar algunas cuantas dimensiones más en torno al Big Data, entre las que se encuentran la volatilidad de los datos y el tiempo, ste último basado en el grado de periodo de espera que deben tener los Big Data. Cada una de estas dimensiones se encuentra mencionada disimuladamente en las que son consideradas características principales, dando por hecho como interactúan.

En el caso de la volatilidad se puede notar más en el volumen y la variedad de los datos, debido a que por los diversos formatos que son almacenados puede complicar demasiado su procesamiento, a pesar de tener herramientas especializadas. Respecto al volumen basta con decir que si no se mantiene un monitoreo constante es posible que los datos sufran un sesgo importante que puedan cambiar la dirección del análisis.

El tiempo es otra de las dimensiones intrínsecas que tienen los grandes datos, que al igual que el anterior se encuentra relacionado con las dimensiones que lo caracterizan. Hay que considerar que el Big Data debe de tener un periodo de maduración para que los datos puedan ser utilizados, al igual que el vino, en donde el añejamiento del producto da como resultado que sea de mucha mayor calidad. En el caso, de las bases de datos cuando se utilizan técnicas como el minado o se cosechan los datos se suele recurrir a lapsos de tiempo medianamente largos para recolectar información suficiente. No obstante, al igual que existen vinos tempranillos con añejamientos de uno o dos años, también hay bases de datos que son analizados en tiempo real, pero, su análisis suele ser un poco más complicado.

**Figura 5. las dimensiones del Big Data**



Fuente: Elaborado y complementado a partir de "El uso del Big Data en el mundo real", IBM, 2012

## 1.5 ¿Qué tipo de información se puede encontrar en Big Data?

A lo largo del capítulo se ha podido observar la composición del Big Data y a su vez, se ha podido vislumbrar un poco la importancia que tiene para la investigación, así como el tipo de datos que se pueden encontrar. Sin embargo, considerando el anterior apartado es necesario mencionar la materia prima con la que el investigador, analista o cualquier persona interesada en el análisis de los paquetes de grandes datos debe conocer, dicha situación se encuentra relacionada con la dimensión de variedad de la que ya explicamos un poco y que en este apartado se dará respuesta a la pregunta ¿qué tipo de información se puede encontrar en Big Data?

Como sabrán, el Big Data tiene una diversidad de formatos o tipos de información que es almacenada por las empresas para analizarla, este gran volumen de datos almacenados, son interpretados como la dimensión de variedad dentro del Big Data que, además de ser extensiones de archivo como docx, xlsx, ptx, cvs, jpg, png, svg, html, css, jar, sólo por mencionar algunos. dentro del documento ¿qué es Big Data? de IBM (2012) se pudieron distinguir cinco categorías que subdividen toda la información proveniente de este océano de datos, la cuales son:

1. Web and social media: esta categoría consiste principalmente en la recolección de datos extraídos de los *buscadores web* o *web browsers* (Google, Firefox, Edge o Yahoo) y las *redes sociales* o *social Networks* (Facebook, Twitter, Whatsapp, Instagram, Snapchat, Youtube). Es decir, toda la información que es subida a redes sociales desde fotografías en Facebook, hasta tweets, además de la información buscada en los navegadores desde un smartphone, Tablet o computador es recolectada y almacenada para ser analizada. Algunas extensiones de archivo que es posible encontrar en esta categoría son los html, css, jar, png, jpg, mpeg, sólo por mencionar algunos de ellos.

2. Machine to machine (M2M): Se engloba prácticamente a todos los dispositivos conectados a internet y/o a otros dispositivos en donde existe una transferencia de datos. Dicha interacción es realizada de máquina a máquina como su nombre lo

indica, es decir, que los datos que son acumulados son extraídos de sensores como: acelerómetro, giroscopio, baumanometro, los sensores de medición de un automóvil, además de las mediciones computarizadas por alguna maquinaria pesada como lo puede ser un alto horno o una prensa hidráulica protocolos, o bien, algo más simple como conectar un teléfono a una red Wi-Fi. El ejemplo más claro para esta categoría es el *internet de las cosas* o *internet of things* que permite la interconexión de dispositivos en una misma red inalámbrica.

3. Big transaction data: Esta es la categoría que menos complicación tiene en cuanto al tipo de datos que suelen ser almacenados, ya que el Big transacción data se encuentra enfocado a todos aquellos registros que estén relacionados con finanzas o que tenga que ver con estados de cuenta, facturaciones, recibos electrónicos. Cabe mencionar que en esta misma categoría son incluidos los Blockchain para minado de criptomonedas, además de la información sobre telecomunicaciones como los registros de llamadas o conversaciones entre dispositivos.

4. Biometrics: En el caso de los biométricos es toda aquella información registrada mediante la digitalización de las partes corporales a través de sensores de proximidad o escáneres entre la información más común que se puede encontrar son el reconocimiento facial, reconocimiento de huellas dactilares, escaneo de retina o reconocimiento de voz, que actualmente se encuentra presente en diversos lugares, desde la seguridad del smartphone que ocupas diariamente, pasando por algo más complicado como lo es una transacción bancaria, hasta algo más complicado como controlar el acceso a un país por medio de escáneres sin necesidad de mostrar tu pasaporte.

5. Human generated: Como su nombre lo indica son todos aquellos datos que son generados de manera consiente por una persona, es decir, un estudiante escribiendo su tesis, un investigador publicando su trabajo en la página de alguna universidad reconocida, pero, además de los periódicos, artículos de revistas digitales, entrevistas, videos, notas de voz y la información guardada para fines de

calidad de un centro de atención telefónica. Es decir, todas aquellas publicaciones que son hechas a través de plataformas digitales.

Si bien es cierto, que es posible confundir alguna de las categorías debido a que los límites de la información no se encuentran totalmente definidos, también es cierto que el investigador deberá tener la pericia para seleccionar la información prudente a su análisis. Ya que, al menos en un primer momento con esta jerarquización establecida por IBM es posible distinguir la utilidad que tiene el Big Data.

A lo largo del primer capítulo se ha podido observar la parte esencial del Big Data, partiendo de la infraestructura física y digital, pasando por la caracterización de sus dimensiones hasta llegar a vislumbrar qué tipo de información puede ser extraída de él. Sin embargo, aún queda por responder ciertas interrogantes que permitirán ver de forma más clara la importancia del análisis de los grandes datos en la investigación y por ende en la utilidad que tiene para la sociología.

## Capítulo 2. La importancia de analizar el Big Data

En este capítulo se pretende explicar la importancia de analizar Big Data a partir de la utilidad que tiene, tomando en cuenta su configuración en el campo de la investigación, además de la presentación de casos en los que fue utilizado el Big Data para situaciones de la vida real. Tomando en cuenta lo que ya se ha leído sobre la composición del Big Data, su importancia puede recaer en tres aspectos esenciales: 1. La información que se desea obtener; 2. El propósito de la investigación; y, por último 3. El beneficio de los datos obtenidos por un análisis de Big Data, referente a la calidad de los datos. (véase figura 6.)

El primer aspecto por considerar es la información que se desea obtener más allá de la técnica de recolección, por lo que se deben tomar en cuenta los siguientes elementos; el primero, destaca la información que se necesita obtener para el análisis, en esta etapa al igual que en sociología se concibe la idea del fenómeno social o lo que se pretenda investigar, considerando la expectativa del investigador sobre la investigación que va a realizar, las posibles fuentes de donde extraerán los datos, para así establecer la viabilidad del proyecto; el segundo, se refiere a la pericia del investigador para ajustar y formular su objeto de estudio. Es decir, aquí ya se plantea el diseño de la investigación formalmente, partiendo de la construcción del estado del arte, la metodología y los referentes teóricos.

Por ejemplo, supongamos que un investigador/sociólogo está planeando hacer un análisis estadístico sobre los factores sociodemográficos en la ciudad de México durante el último siglo para comparar la extensión de la mancha poblacional. Por otro lado, algún colega pretende hacer un estudio un poco más abstracto sobre el reconocimiento de patrones socioculturales en redes sociales. En ambos casos, se planea hacer un análisis de Big Data, sin embargo por las condiciones de cada estudio el análisis deberá ser realizado en diferentes niveles, mientras que en el primer caso es un análisis estadístico correspondiente a un estudio de corte cuantitativo, resulta mucho más sencillo recurrir a bases de datos estructuradas como las de los centros especializados (INEGI o CONEVAL, entre otros) recurriendo

a un nivel superficial de Big Data como el Open Data; el segundo se presenta más como de corte cualitativo debido a que sus variables no se encuentran tan visibles y es posible que deba crear nuevas categorías, por lo que es recomendable realizar un análisis de redes sociales (ARS) complementándolos con otras técnicas tradicionales de la sociología.

El segundo aspecto por considerar es el propósito de la investigación o siendo más específico es el propósito de realizar el análisis de Big Data que al igual que con el aspecto anterior contiene tres elementos que ayudan a su descripción; el primero es la utilidad que tiene el análisis, independientemente de los datos generados, es decir, el sociólogo -hablando de manera particular- debe concebir la utilidad que tiene realizar un análisis de Big Data dentro de su proyecto, por lo que dicha justificación se encuentra definida en el diseño de investigación.; el segundo factor, se refiere a la aplicación que tendrán los datos al concluir el estudio, es decir una vez definida la utilidad del Big Data en la investigación, debemos determinar para qué serán usados los datos que arrojará el análisis; el tercer factor que se debe considerar son las limitaciones del campo de estudio, por lo que el investigador debe tener en cuenta cuales son las circunstancias y las restricciones que tiene la sociología en el análisis de macrodatos respecto al tema tratado. Por ejemplo, un investigador/sociólogo después de haber recolectado información durante un periodo de seis meses, tiene en mente realizar la configuración de los escenarios más probables dentro de un momento coyuntural, digamos una elección, pero además de eso, con los datos que ya recolectó pretende revelar patrones de comportamiento electoral para conocer la probabilidad de cambio en la tendencia electoral.

Por último, el tercer aspecto trata sobre el beneficio de los datos obtenidos por un análisis de Big Data, que a diferencia de los anteriores este reduce todos sus elementos a uno sólo, la calidad de los datos. Si bien, este último elemento entra en sintonía con los anteriores, debido a que además de depender del campo de estudio de donde se presente el análisis y las aplicaciones de los datos dentro de ese sector,

permitiría al investigador/ sociólogo fundamentar de forma plena la importancia de analizar los grandes paquetes de información.

Ahora bien, por calidad de datos se refiere al control preservado de estos en la etapa de integración durante el proceso de análisis, asegurándose de que “reúne entre sus atributos la exactitud, completitud, integridad, actualización, coherencia, relevancia, accesibilidad y confiabilidad necesarias para resultar útiles al procesamiento, análisis y cualquier otro fin que un usuario quiera darles” (powerdata, 2013). En otras palabras, la calidad de los datos es el proceso de verificación de los datos respecto a la fiabilidad y la confianza como se indica en las dimensiones existentes del Big Data. Sumado a lo anterior, el control en la calidad de los datos es ni más ni menos que la gestión de la información.

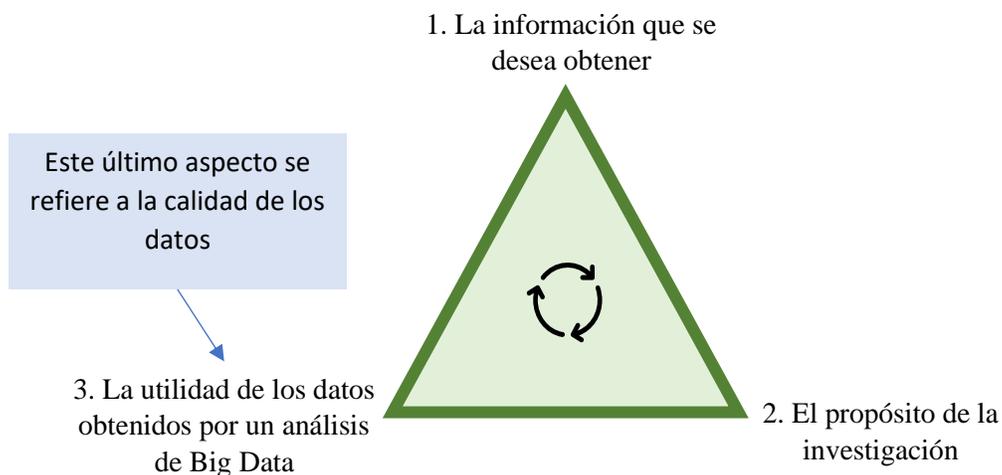
Cabe mencionar que la practicidad que tiene el Big Data dentro de la investigación resulta enriquecedora para la sociología, por lo que de esa forma se perfeccionaría la labor del sociólogo al tener un campo de estudio más amplio y sin limitarlo a sólo las variables tradicionales, complementando además sus ejes de análisis y construyendo nuevas categorías, que le permitan que los resultados obtenidos por Big Data dependan más de la calidad de la información que pueda ser recolectada mas no de la cantidad de información que pueda ser procesada, siendo un elemento versátil por las bondades que ofrece en estudios de corte cualitativo o cuantitativo.

Aunque pareciera relativo indicar la importancia del Big Data con aspectos que pecan de ser subjetivos, en realidad “el uso de Big Data ha ayudado a los investigadores a descubrir cosas que les podrían haber tomado años en descubrir por sí mismos sin el uso de estas herramientas, debido a la velocidad del análisis, es posible que el analista de datos pueda cambiar sus ideas basándose en el resultado obtenido y re trabajar el procedimiento una y otra vez hasta encontrar el verdadero valor al que se está tratando de llegar” (IBM,2012. 9). En otras palabras, la importancia que adquiere en la investigación yace en los datos que se pueden encontrar dentro de ese océano digital, partiendo de la habilidad del investigador para reinterpretar sus variables, formulando nuevos parámetros más específicos que se adecuen a su análisis.

## 2.1 El Big Data en la investigación

La mayoría de las ocasiones cuando se habla de Big Data sólo se suele abordar la parte conceptual de caracterización. Sin embargo, poco se habla de la frecuencia con la que podemos encontrar el Big Data aplicado en el mundo real, existen muchos ejemplos donde pueden ser visibles estas aplicaciones, desde el momento en el que abrimos una app de geolocalización en nuestro smartphone hasta cuando realizamos la cotización del seguro para un vehículo, en cada una de estas situaciones de manera consciente o inconsciente el usuario contribuye a la alimentación de un Big Data debido a que termina brindando parte de su información personal; en el caso de los servicios de geolocalización el usuario conscientemente comparte su ubicación aproximada con otro, mientras que en el segundo caso parte de la información que ofrece el usuario es enviada a los bancos de datos de cada empresa para determinar la factibilidad que tiene esa persona de adquirir dicho seguro.

**Figura 6. Tres aspectos determinan la importancia del análisis de Big Data**



Fuente: Elaboración propia

La incorporación del Big Data dentro del campo de investigación ha sido un determinante coyuntural debido a que "... esta explosión de "grandes datos" está transformando la manera en que se conduce una investigación adquiriendo habilidades en el uso de Big Data para resolver problemas complejos relacionados con el descubrimiento científico, investigación ambiental y biomédica, educación, salud, seguridad nacional, entre otros" (IBM, 2012, 8). El Big Data dentro de la investigación se ha vuelto un pilar para las empresas que buscan soluciones rápidas para situaciones que tengan que ver con el mercado, gracias a que vieron un alto potencial en la explotación de los datos personales con la finalidad de aumentar la venta de sus productos. Por lo que, tras visualizar ese sector de interés, de alguna forma lograron crear una herramienta con gran flexibilidad y practicidad, ya que evaluar clientes con la intención de conocer qué es lo que buscaban de un producto o servicio específico con técnicas como el minado de datos permitieron captar una mayor cantidad de opiniones sobre sus productos.

Este cambio de perspectiva en la forma de resolver los problemas empresariales no tardó en replicarse en muy poco tiempo por otros sectores ya que "... el análisis con Big Data puede aportar al conocimiento de fenómenos sociales complejos para mejorar la toma de decisiones" (Meneses, 2018: 419). Por ello, las empresas que comenzaron a apostar por este tipo de análisis de datos para aumentar en gran medida su mercado con la utilización de técnicas propias en la construcción de bases de datos que ha resultado en una herramienta útil para la investigación proveyendo una mayor cantidad de datos y, por lo tanto, los estudios que son realizados pueden arrojar mayores indicadores que resuelvan alguna problemática. "... el diseño tecnológico aparece como una forma de opacidad de las redes sociales digitales que, bajo el halo de la sociabilidad, son en realidad empresas colectoras de datos" (Meneses, 2018: 426). En este punto, el proceso de recolección se encuentra mucho más notorio hoy en día con empresas que ofrecen servicios en streaming como YouTube, Netflix, Amazon prime video, en donde la implementación de algoritmos incentiva la interacción del usuario con la plataforma

para aumentar su consumo. Es decir, entre más contenido consumes, más fácil las empresas harán target en tu huella digital para ofrecerte sus productos.

Sumado a lo anterior, la migración de los medios de comunicación tradicionales hacia plataformas web, la digitalización de contenido y la popularización de las redes sociales ha dado como resultado la apertura de nuevos campos de estudio, debido a que hoy en día al encontrarse normalizadas actividades como la publicación de fotos o videos en redes sociales, compartir ubicaciones e inclusive valorar la calidad de los servicios, ha derivado en la transformación de una sociedad análoga a una digital haciendo que las prácticas sociales cambien a modo de que la misma sociedad establezca sus funciones a través de tendencias como está sucediendo actualmente con el uso de las redes sociales. Haciendo alusión a lo que mencionaba Luhmann con los medios electrónicos como un medio de difusión a través de computadoras, describiendo los impactos significativos producidos por estos, como las limitaciones y la transformación del proceso comunicativo respecto al paradigma tradicional del emisor – receptor.

No obstante, además de la adaptación digital de los medios de comunicación, “...la recolección de datos y las múltiples transacciones que realizan empresas como Google, Facebook, Amazon, Apple, así como las compañías de telecomunicaciones —en algunos casos—, no se llevan a cabo mediante contratos transparentes sobre su utilización y protección. Esta situación es observable en las redes sociales digitales, cuyos usuarios aceptan condiciones de uso de sus datos sin conocer a cabalidad lo que implican. Entre desconocimiento, tecnicismos y prácticas opacas se aceptan condiciones sin conocer, en la mayoría de los casos, las consecuencias” (Meneses, 2018: 426). Provocando que sea revelado otro punto de vista respecto a la privacidad de los datos y con ello los usuarios de esas plataformas de alguna forma ha comenzado una resignificación al termino convencional *de privacidad de la información* una vez que logran iniciar sesión en internet, situación que puede interesar al sociólogo en plena era digital.

A pesar de que las empresas ya se encuentren trabajando con este tipo de análisis, existen sectores que aún se encuentran descubriendo el mundo del Big Data como son las farmacéuticas que utilizan esta herramienta para verificar y comparar las composiciones químicas de las vacunas, o el sector médico con la explotación de los Data Lake haciendo posible la creación de repositorios de información que calculan la probabilidad de los pacientes a desarrollar enfermedades como el cáncer, pero a su vez también tienen la posibilidad de diseñar tratamientos mucho más efectivos para esas enfermedades.

“En tanto que la práctica científica ha sido ya modificada con el análisis de cuantiosos datos en diversos campos disciplinarios, como la física y la biología” (Heiberger y Riebling, 2016; Meneses, 2018: 419). La presencia del Big Data en la investigación ha potenciado la posibilidad de reinterpretar fenómenos sociales debido a la información recolectada de múltiples fuentes, que al igual que las empresas lo usan de una forma estratégica para atraer un sector de clientes potenciales “... poco a poco las huellas humanas embebidas en satélites scanners y redes sociales digitales comienzan a ser cuantificadas y analizadas por científicos sociales” (Meneses, 2018: 419). Es allí en donde la sociología puede verse involucrada, ya que la presencia del Big Data vino a revolucionar la manera en que se pueden abordar los fenómenos sociales con herramientas actuales y complementadas con las tradicionales de la sociología.

La incorporación del Big Data a la investigación ha sido de gran ayuda para el sector académico y empresarial teniendo en cuenta que “... los datos no existen fuera de las ideas, de los instrumentos, las prácticas y el contexto que enmarca su creación e interpretación” (Gitelman, 2013; Meneses 2018: 424). Tomando en cuenta el uso de su propia infraestructura para analizar la curva de crecimiento de un negocio o para establecer sectores potenciales de expansión, es algo que las empresas nunca se imaginaron tener en sus manos para mejorar sus productos o servicios y mucho menos que esta herramienta fuera a ser una de las más utilizadas no sólo por las empresas sino también por el sector académico.

## **2.2 Los casos de uso del Big Data (casos prácticos del Big Data en el mundo real)**

Es cierto que el Big Data es una de las herramientas más flexibles en cuanto a la versatilidad en sus aplicaciones y a su vez, una de las más complejas por la volatilidad que llegan a tener sus componentes, por lo que difícilmente solemos percatarnos de cuando somos parte del análisis de este monstruo informático. Por ello, la intención de este apartado no es sumergir al lector en la revelación de alguna teoría de conspiración que provoque que deje de utilizar sus dispositivos, sino exponer algunos de los casos en los que el Big Data ha sido aplicado.

Lo primero que se viene a la mente cuando una persona habla sobre las aplicaciones del Big Data en el mundo real, son millones de líneas de código conectando todo, y aunque en parte tienen razón que los grandes datos se componen de la interacción de las computadoras con el usuario, también es cierto que en todo momento se están realizando análisis de Big Data y nosotros como usuarios de la red estamos siendo parte de ellos, es decir, nuestra huella digital en la World Wide Web está siendo monitoreada todo el tiempo y, por tanto, estamos sujetos a la publicidad en internet. Esto se debe principalmente a que áreas como el marketing, la publicidad, la política, la gestión de empresas, el periodismo, la ingeniería en software, la medicina, solo por mencionar algunas, han comenzado a adoptar la utilización del Big Data dentro de sus investigaciones, ya que el manejo de grandes paquetes de información permite un mayor alcance.<sup>9</sup>

### **Marketing/publicidad**

Este monitoreo por medio de cookies<sup>10</sup> que muestra sugerencias de productos en la pantalla principal del ordenador mientras estamos navegando, no es la única

---

<sup>9</sup> Como una recomendación, si en algún momento tienen la oportunidad, abran su navegador de confianza y hagan la cotización de algún dispositivo electrónico, seguro para automóvil/motocicleta o cualquier producto, notarán que al cabo de un par de horas empezarán a ver anuncios de la tienda/compañía donde hicieron su cotización.

<sup>10</sup> Se define como cookies a los paquetes de datos que son almacenados en los dispositivos de forma automática a través de un navegador web cada vez que visita algún sitio web. Es decir, es enviada desde los servidores de la página/dominio web hacia el usuario, similar a una hoja de registro de visitas.

aplicación que tiene el Big Data, si bien es una de las más populares y que realizan diversos sitios web para ofrecer sus servicios, la diversidad de aplicaciones que tiene el Big Data permite sacar más provecho a partir de una visita, como se ve reflejado en el marketing, siendo ésta una de las áreas en las que más se ha desarrollado el Big Data, principalmente por la alianza que tiene con las empresas siendo pioneros en la aplicación de Big Data de forma comercial con la intención de evaluar el mercado. Es decir, la utilización de grandes paquetes de información para analizar su sector mercantil, promoviendo la evaluación de productos o servicios para plantear escenarios en los que alguna marca pueda expandirse y aunque el ejercicio anterior es un ejemplo de la focalización hecha por las empresas a través de los grandes datos, otro de los ejemplos que más destacan es con la utilización de los servicios de transporte por aplicación (UBER, Cabify, Didi, etc.), que si han tenido la oportunidad de usar alguna de estas, se habrán percatado que al concluir el viaje usted como usuario tiene la posibilidad de evaluar el servicio, calificando al personal, o bien haciendo observaciones sobre las condiciones del transporte. Bien, esto sucede principalmente por varios motivos, el primero es porque la aplicación recopila la experiencia del usuario utilizando el servicio; el segundo es que además de captar la experiencia del usuario, evalúa la eficiencia de sus empleados y, por último; el tercero se refiere a la probabilidad de volver a usar el servicio. Pero no tan sólo la utilización de estos bloques de información permite visualizar la experiencia del usuario con la plataforma con el fin de evaluarla, sino que en conjunto la experiencia obtenida del pasajero permite que la empresa conozca el nivel de satisfacción total del usuario, para así poder hacer modificaciones o actualizaciones que mejoren la aplicación y con ello tener un mayor alcance dentro del mercado actual.

El marketing aplicado al Big Data evalúa propiamente la retención y la pérdida de clientes a través de modelos predictivos diseñados con Big Data; ahora bien, desde el punto de vista de la gestión empresarial evalúan el riesgo de inversión a través de algoritmos o patrones de comportamiento diseñados para estimar las ganancias y las pérdidas que pueda tener una empresa; asimismo, los bancos también suelen usar este análisis, utilizando la geolocalización del cuentahabiente, la fecha y hora

en que es realizada una transacción, además de la plataforma en que la realizó. Esto, para prevenir en la mayoría de las ocasiones las transacciones fraudulentas.

## **Política**

Cabe mencionar que la recolección y análisis de datos de forma periódica sobre plataformas digitales no sólo se quedó en la parte comercial, sino que su implementación también ha sido dirigida hacia la esfera política. La incorporación del Big Data como una herramienta de análisis en la política se ha llevado a cabo con la evolución de la tecnología, es decir, se ha ido adaptando conforme a las necesidades de los propios usuarios destinando recursos en el diseño de campañas electorales, la evaluación de políticas públicas y el seguimiento en temas que conforman la agenda legislativa. En otras palabras, una estrategia política basada en el cambio variable complementado de la opinión pública con la información extraída del seguimiento de redes sociales y buscadores web, con el fin de desarrollar nuevas estrategias que establezcan ventaja sobre el otro -en el caso de las contiendas electorales- o bien que permitan vislumbrar sectores vulnerables que ayuden al diseño de una política pública. En ambos casos existen ejemplos de cómo se ha visto involucrado el Big Data dentro de esa esfera.

Algunos de los ejemplos con mayor relevancia de la utilización del Big Data en la política han sido aquellos productos resultados principalmente de la utilización de las redes sociales. Tal es el caso de *Cambridge Analytica* en el que se vio involucrada la red social *Facebook* y la consultoría con el mismo nombre durante el 2016, en donde se dio a conocer que a través de un *Quiz* dentro de la red social se crearon nodos de información con los que era posible perfilar la postura política de los usuarios, de esta forma accediendo a los datos de aproximadamente 87 millones de personas. Si bien, el cabildeo en las redes sociales es algo que se acostumbra hacer en la actualidad durante el periodo electoral, recordemos que Cambridge Analytica era una consultoría política especializada en hacer análisis de redes sociales y Big Data enfocados particularmente a temas coyunturales, y que para ese momento tenía la tarea de perfilar a sujetos potenciales para después bombardearlos con anuncios, imágenes negativas, cifras maquilladas, noticias

falsas o cualquier información que permitiese cambiar la tendencia del voto en la contienda electoral. Sumado a eso, *Facebook*, la red social con mayor número de usuarios era el medio necesario para cumplir con su labor, por lo que utilizar grandes paquetes de datos extraídos de esa red no sonaba tan descabellado. Aunado a lo anterior, en el 2016 las políticas de privacidad de *Facebook* sobre aplicaciones de terceros dentro de la plataforma no se encontraban reguladas como hoy en día. Cabe señalar que esa misma estrategia realizada en 2016 por el equipo de asesores de Donald Trump ya había sido puesta a prueba tiempo atrás, en 2012 durante la campaña de reelección de Barack Obama, su equipo de trabajo recurrió a un análisis de redes sociales y Big Data solicitando a la población que compartieran información que pudiera vislumbrar las inquietudes y necesidades que ayudasen al diseño de su campaña.

Pese a toda la controversia en la que se ha visto implicada Cambridge Analytica en cuanto a la seguridad de la información, también existen otras aplicaciones dentro de este sector como en el caso de México, que recurrentemente se diseñan campañas electorales con una infraestructura de redes sociales, debido a que actualmente el 65.8%<sup>11</sup> de la población se encuentra conectada a internet a través de algún dispositivo (PC, Tablet o móvil) indicando que es una de las formas más factibles de llegar a los electores, ejemplo de ello, fueron las pasadas elecciones presidenciales en donde las plataformas más utilizadas principalmente para impulsar las campañas fueron Twitter y YouTube, complementado claro está con algoritmos en Google que hicieran un match de palabras que orientaran a los votantes a las páginas web de los candidatos. Es decir, sí introducían <<candidatos de las elecciones presidenciales en México 2018 y/o candidatos presidenciales 2018>> en el navegador, al momento de presionar *enter* y consultar los resultados, el primero que salía era la página de José Antonio Meade (en ese momento candidato del Partido Revolucionario Institucional), y lo mismo ocurría con cualquier búsqueda que incluyera las palabras <<candidato/ elecciones/ presidenciales/ 2018>> esto se debe a los algoritmos de *matching words* en las plataformas y del

---

<sup>11</sup> Cifra indicada por la asociación de internet México hasta la fecha de mayo del 2019.

cual se hablará más adelante en este capítulo. Asimismo, se han dado a conocer casos como el de Victory Lab, una empresa cuya labor principal es posicionar hashtags en redes sociales y difundir noticias falsas con la utilización de herramientas de Big Data.

### **Periodismo de datos**

Por otro lado, el periodismo también ha sacado ventaja de la utilización del Big Data, ya que gracias a él han surgido nuevas herramientas de utilidad para desarrollar de mejor forma la labor de los periodistas, pudiendo ampliar su información y consultar fuentes alternas que complementen sus historias, así como conseguir documentos que las fundamenten. Con esto último, tal vez recordarán a *wikileaks*, un sitio web creado en 2005 por Julian Assange en donde se publicaban documentos clasificados de índole gubernamental, del que se desprendieron diversas investigaciones acerca de la corrupción de los gobiernos, espionaje gubernamental y operaciones militares secretas, que fueron publicadas por diversos medios de comunicación como The New Yorker, The Guardian, The New York Times, The Wall Street Journal y The Washington Post, sólo por mencionar algunos. Así también como los *Panama Papers*, una investigación realizada por Gerald Ryle con la que se expusieron los paraísos fiscales alrededor del mundo viéndose involucrados principalmente políticos y empresarios. Estos casos son una prueba del alcance que tiene el Big Data que, aunque pueden o no ser considerados como producto directo de este, sin duda alguna fungen como una fuente de información y dan pauta al periodismo de datos, una línea de investigación vinculada directamente al Big Data.

### **Medicina**

Dicho sea de paso, el campo de la medicina y de la salud se ha visto beneficiado con las aplicaciones del Big Data, debido a que han contribuido con aportes significativos en sus distintos campos con la creación de repositorios de información centralizados que almacenan datos sobre distintos tipos de enfermedades como cardiovasculares, crónicas degenerativas, cáncer, además de identificar nuevas cepas de virus. “Archivar centralmente datos de pacientes con el fin de encontrar

nuevas interrelaciones ya se está convirtiendo en realidad en Estonia y en Reino Unido, y otros países también están trabajando para impulsar el desarrollo de la medicina digital, dentro del concepto más amplio de salud digital” (Bayer, 2017). Haciendo posible que con la información recolectada se puedan desarrollar nuevos tratamientos contra enfermedades e incluso “... el análisis de grandes volúmenes de datos puede ayudar a frenar una epidemia como el ébola al mapear el recorrido de los casos, mediante teléfonos celulares, alertando a los servicios de salud de los diversos continentes” (Meneses, 2018: 419). Ya que se puede tener un mejor control sobre patógenos externos y actuar inclusive con la localización del paciente cero antes de que estos puedan diezmar a la población.

Además de la creación de repositorios de información que permiten analizar y diagnosticar enfermedades de una manera mucho más efectiva para el desarrollo de nuevos tratamientos y el mapeo del genoma humano, “El big data puede ser eficaz a la hora de generar iniciativas que ayudan a gestionar mejor los negocios y responder de una forma más efectiva a las demandas de los pacientes. Así, gracias a los macrodatos, los farmacéuticos pueden recopilar datos de los colegios de farmacéuticos, conocer el número de visitas concretas que reciben en su establecimiento, identificar las necesidades de los pacientes o gestionar mejor el stock de productos y los precios” (Pfizer, 2016). Es decir, una mayor eficiencia en el sistema de salud, dando prioridad de manera eficaz a los pacientes y a su vez fomentar la medicina preventiva.

Ahora bien, en los últimos cinco años, la medicina preventiva de la mano con la tecnología ha cobrado fuerza, actualmente, es posible monitorear nuestra salud con los dispositivos que usamos diariamente como *teléfonos* o *relojes inteligentes*, proveyendo un mejor control en la salud debido a que en ellos se incorporan sensores que miden el ritmo cardiaco, la presión arterial, el oxígeno en la sangre e inclusive se cuentan el número de pasos que damos al día, siendo posible tener un registro preciso sobre nuestra actividad, de esta forma los médicos puedan visualizar mejor nuestro estado de salud para el tratamiento adecuado. Cabe mencionar que empresas como Bayer y Pfizer ya han incorporado el análisis de Big

Data dentro del sector farmacéutico, apoyándose principalmente con la inteligencia de datos. Por ello, uno de los pioneros en llevar a cabo la incorporación del Big data al sector salud es España, con el Hospital La Fe en Valencia que en coordinación de las instituciones de salud de ese país desarrolló una plataforma que "...permite combinar datos masivos tradicionales con datos individuales y poblacionales, permitiendo organizar la información de manera más efectiva" (Pfizer, 2016). Esto, con la finalidad de generar mejor información para abordar de manera eficiente la caracterización de enfermedades y la administración de los servicios de salud. En resumen, el uso de herramientas como el Big Data en el sector salud ha abierto nuevas posibilidades para el diagnóstico y tratamiento de enfermedades, además de ampliar el panorama sobre la incipiente realidad de los problemas de salud que aquejan a la población, sumado a la posibilidad que hay de su detección y tratamiento en tiempo real.

### **Match de palabras y algoritmos de recomendación es igual a una herramienta de rastreo de información**

De la misma forma, como sucede con el ejemplo anterior sobre los servicios de salud, la simple actividad de navegar en internet, poseer un perfil en las redes sociales e inclusive tener contratado servicios en streaming como Spotify o Netflix, contribuyen a la creación de bancos de datos que llevan a cabo un registro de nuestro comportamiento en internet con el fin de sacar provecho de nuestra actividad. Tal es el caso de los pareos de palabras o *matching words*, que detectan palabras específicas por medio de algoritmos de reconocimiento con la finalidad de mostrar la mayor cantidad de información relacionada a ella. Actualmente los dispositivos que usamos tienen la capacidad de escucharnos para realizar alguna acción en el dispositivo como reproducir música o leer mensajes, pero también son capaces de determinar nuestros gustos y realizar sugerencias acerca de ellos.<sup>12</sup>

---

<sup>12</sup> Para más información véase en Forbes (2019) "Google admite que escucha conversaciones con su asistente virtual" <https://bit.ly/2Qp0Fg4>

Esto se debe a que el *matching word* o pareo de palabras funciona mediante inteligencia artificial, algoritmos de machine learning y repositorios de información sustraídos de Big Data. Haciendo posible que exista una focalización de contenido como la que se lleva a cabo dentro de las redes sociales, siendo en cierta forma capaces de determinar los gustos del usuario acorde a su ciber-personalidad, como se vio con Cambridge Analytica, sólo por mencionar un ejemplo, y más aun con los servicios digitales como el de Spotify o Netflix, en donde podemos observar "...cómo los gustos de sus usuarios convertidos en datos se traducen en una rigurosa personalización que le permite a la empresa satisfacer preferencias personales, lo cual supera en mucho a las mediciones de ratings de la era análoga" (Meneses, 2018: 424). Curiosamente, el pareo de palabras y la focalización de contenido, si bien, son herramientas útiles para seleccionar contenido adecuado a una persona, es también utilizado para rastrear información específica, debido al vínculo que tiene con las masivas cantidades de información que emanan de los repositorios de las empresas.

Aunque hoy en día el rastreo de información se ha vuelto una tarea en cierta medida mucho más simple, debido a gran cantidad de dispositivos conectados en la red, rastrear información aún puede ser una tarea exhaustiva hasta para el más ávido investigador, ya que diariamente son generados nuevos datos y mantener un seguimiento constante de ellos puede ser complicado, por lo que se requiere la implementación de herramientas que faciliten nuestra tarea en la búsqueda y recolección de datos, he ahí que el pareo de palabras funge como una de las aplicaciones que tiene el Big Data, dando la posibilidad de realizar seguimientos de manera automatizada o semiautomatizada a través de una combinación de machine learning e inteligencia artificial. Ejemplo de ello, es Google Alert, cuya finalidad es recolectar información sobre temas, palabras, textos o personas, ya que el usuario únicamente coloca el tema de interés en un cuadro de diálogo y la plataforma arrojará todos los resultados que se obtengan sobre ese tema. Ahora bien, al ser una herramienta relativamente sencilla contemplada para rastrear información, los resultados de nuestra búsqueda son compilados en un *feedback* enviado a una dirección de correo electrónico previamente solicitada que funciona como punto de

control para el usuario. Dicho sea de paso, Google Alert es una clara representación del Big Data en su nivel más básico, es decir que deja al usuario alejado de las líneas de código y del conocimiento de lenguajes de programación. Cabe mencionar que plataformas como Google Alert son utilizadas principalmente para realizar seguimientos de temas coyunturales o tener una mayor cobertura en los medios sobre temas de interés público como tratados internacionales, migración, medio ambiente, violencia, inclusive economía global, sólo por mencionar algunos ejemplos.

En este punto, el lector habrá notado similitudes en los casos que se ha hecho uso del Big Data y la importancia respecto a este tipo de herramientas en la investigación. Obviando la necesidad del volumen, la velocidad y la variedad de fuentes que componen al Big Data, la mayoría de los casos en donde se ha utilizado han sido relacionados con temas que le competen a la sociología como lo pueden ser aspectos sociodemográficos, ideológicos, bienestar social e inclusive económicos. Por lo que recurrir al Big Data puede permitir un avance significativo en cuanto a análisis de datos, visto desde una perspectiva sociológica, el uso de los grandes datos puede contribuir a la apertura de nuevos campos de investigación que anteriormente no eran observables, y en los que la sociología debido a su formación tradicionalista, no ha podido penetrar, dado que el estudio de un fenómeno científico o social depende de la manera en que el investigador aborde dicho fenómeno, además de la fuente de información que utilice para sustentar sus observaciones.

### **2.3 Los niveles de profundidad de Big Data**

Como se habrán dado cuenta en el apartado anterior, el uso de Big Data puede ser diverso, debido a la aplicación multidisciplinar que tiene gracias a la selección de múltiples fuentes que alimentan al océano de información y con ello a los grandes datos. Ahora bien, recordaran que en el primer capítulo se mencionan los tipos de información que circulan en internet, estos se pueden dividir en las búsquedas, los

datos financieros, la comunicación entre máquinas, los datos provenientes de sensores y los datos generados por conocimiento humano, dando por sentado que ya por el nombre en sí podemos especular de qué se trata cada uno. Sin embargo, cuando se habla de la complejidad del Big Data, va más allá de la interacción de sus componentes básicos, encontrando gran parte de ella en la integración de la información, es decir, las múltiples fuentes y los diferentes formatos han hecho que se vuelva complicado discernir la información que puede ser accesible para los usuarios. Por ello, existe la necesidad de contemplar los niveles de profundidad que tiene el Big Data que ayudarán en principio para desmenuzar la complejidad que puede tener el análisis de los grandes datos.

Dicho lo anterior, este apartado tiene la finalidad de explicar los distintos niveles de profundidad que tiene el Big Data, ya que en ellos se puede determinar la viabilidad que tiene el análisis de Big Data para la sociología considerando que entre más se escale el nivel de profundidad, mayor será la complejidad en el proceso de análisis e investigación, debido a que las herramientas que permiten al usuario la visualización de los datos se irán rebuscando.

### **2.3.1 Open Data**

Comenzando con el primer nivel en Big Data se localiza el Open Data, el cual hace referencia a todos aquellos datos a los que nosotros como usuarios de plataformas digitales podemos acceder y hacer uso de ellos sin restricción alguna. Si bien el Open data surge como un movimiento para la desclasificación de datos gubernamentales para todos aquellos que estuvieran interesados en consultarlos, realmente sus inicios datan de la transformación de la World Wide Web y la digitalización a todo lo que actualmente conocemos como medios electrónicos. Ahora bien, al ser un modelo de transparencia de la información resultado de movimientos de apertura de datos, se debe tomar en cuenta que la caracterización que se pretende dar aquí sobre el Open Data o datos abiertos es como una unidad de información desclasificada que se encuentra integrada a los bloques de datos que componen Big Data, por lo que puede haber diferencias en el tratamiento del concepto respecto a su caracterización original.

La apertura de los datos, la evolución de internet con la web 2.0 y los registros de dominios electrónicos dieron como resultado la creación de bloques de información tan extensos que se convirtieron en uno de los precursores en la construcción de los Big Data, debido a que en los inicios de la Word Wide Web se pretendía la interconexión de ordenadores para consultar, transferir e incluso se pudiese analizar la información disponible en ellos, "...en las últimas décadas comenzó a hacerse visible un cambio radical en la forma de comunicarse, en los medios utilizados para intercambiar y procesar la información, en la manera de difundir la producción intelectual, y en el consumo de las bases de datos a través del uso intensivo del fenómeno llamado Internet." (Letteruci en Bron, 2015:17). Ya que actualmente la mayoría de los dispositivos se encuentran conectados a las redes inalámbricas, por lo que "...Los datos nos invaden en todos los contextos de la sociedad, generados por las aplicaciones de la web 2.0, los gobiernos, las investigaciones científicas, las publicaciones o las organizaciones" (Ferrer-Sapena, Sánchez-Pérez, 2013: 150). Permitiendo así que el acceso a la información, la publicación de datos en las redes y la consulta de los medios digitales funjan como una fuente potencial para el Big Data, de allí la importancia que se tiene contemplarlo como parte de su estructura de niveles.

Pero ¿qué es el Open Data? "Se entiende como open data la apertura (o puesta a disposición de cualquiera) de datos digitales, pero en este concepto también tiene cabida la de datos en documentos impresos" (Ferrer-Sapena, Sánchez-Pérez, 2013: 150). Es decir, una enorme cantidad de datos depositados en archivos y bibliotecas digitales y sin digitalizar, de los cuales son considerados como parte del dominio público. Si bien, el término Open Data se refiere a la apertura de la información, "... lo que hoy conocemos como Datos Abiertos son en general fuentes de datos que históricamente han sido cerrados y que han estado bajo control de organizaciones -públicas o privadas- y cuyo acceso ha estado restringido mediante diferentes tipos de limitaciones, licencias, derechos de autor y/o patentes" (Bron, 2015: 10). Es decir, cuando se habla sobre los datos abiertos disponibles son en su mayoría datos geográficos, estadísticos, médicos, biológicos, bursátiles e históricos,

incluyendo también a todas las producciones científicas y académicas ya publicadas.

El Open Data se podría considerar como el primer nivel de profundidad que tiene Big Data, por la interacción que tiene con el usuario, debido a que los datos que allí circulan "...pueden ser utilizados libremente, reutilizados y redistribuidos por cualquier persona, permitiendo la creación de obras derivadas que reconozcan la misma licencia" (Ferrer-Sapena, Sánchez-Pérez, 2013: 151). Ejemplo de ello son los proyectos de código abierto, sin embargo, "el concepto open data va más allá de la mera reutilización. Su filosofía pretende que los datos se encuentren disponibles de manera libre para todo el mundo, sin restricciones de copyright" (Ferrer-Sapena, Sánchez-Pérez, 2013: 150). De allí que existen diversos sitios web que proveen a los usuarios de contenido gratuito o una parte de él, claro ejemplo son los portales de las instituciones públicas que permiten consultar información, o bien, páginas web que distribuyen contenido digital de forma gratuita, añadiendo también a los foros de programación. No obstante, más allá de la representación de una filosofía de transparencia en la información, el Open Data representa la interacción de las múltiples fuentes de datos encausadas en la infraestructura del Big Data.

Ahora bien, antes de continuar es necesario aclarar que como en cualquier otro concepto, el Open Data también posee características que ayudan a definirlo. Dichas características presentadas en *Open data, big data: ¿hacia dónde nos dirigimos?*<sup>13</sup> Son definidas en tres principales; 1. La disponibilidad y acceso en la que se refiere a que los datos deben estar disponibles para su descarga y modificación a través de internet; 2. La reutilización y redistribución considerada como las condiciones óptimas para la utilización de los datos, pudiendo ser combinados con otros paquetes de información previa o posteriormente descargada; 3. La partición universal en la que se destaca que todos los usuarios

---

<sup>13</sup> Texto publicado por Antonia Ferrer Sapena y Enrique A. Sánchez Pérez en el Anuario ThinkEPI 2013, en el cual desmitifican la utilización del Open Data.

no deben tener restricciones para la utilización, reutilización o distribución de los datos.

Acorde a esa caracterización, el open data ha dado como resultado una mejor y mayor producción de datos referentes al aspecto social. Ya que “Los Estados, producto de su función recopilan y producen grandes volúmenes de datos de todo tipo (climatológico, económico, social, cultural y muchos otros), los cuales generalmente se encuentran con accesos restringidos, en muchos casos ni siquiera se conoce su existencia, en formatos no estandarizados y con una gestión bastante deficiente por parte del Estado” (Barros, 2011: 26). Dicho de otro modo, las instituciones se han encargado de construir indicadores que ayudan no tan solo a la sociología para estudiar el comportamiento de la sociedad, ejemplo de ello, es INEGI con la medición y las estimaciones del crecimiento poblacional y CONEVAL con las estimaciones de los indicadores socioeconómicos. Aplicándolo de manera particular en la sociología, el Open Data se encuentra presente durante toda la investigación, desde la construcción del diseño de investigación hasta su conclusión. Inclusive ahora mismo es posible notar la interacción que tiene el Open Data con la construcción de este trabajo de investigación.

Por otra parte, los datos que fungen como indicadores o índices “... al ser recogidos de entidades prestigiosas tienen sello de calidad y fiabilidad, posibilita que los usuarios accedan y reutilicen los datos de manera libre y gratuita...” (Letteruci en Bron, 2015:17). Lo que satisface las dimensiones del valor y la veracidad que forman parte del Big Data, ya que al ser paquetes de información extensos salidos de instituciones son verificables, lo cual hace que aumente su valor como dato de interés. Sumado a lo anterior, la masificación de la información dentro del océano digital ha logrado que los usuarios de la world wide web puedan tener acceso a la información gubernamental a través de los portales de transparencia, ejemplo de ello, es el Instituto Nacional de Transparencia, Acceso a la información y Protección de Datos Personales (INAI). En el caso de México, que si bien, promueve la consulta y el acceso a la información para cualquier persona que tenga interés en los datos de cualquier institución, es una clara representación sobre la interacción que tiene

open data debido a que al ser un movimiento que promueve la liberación de datos no textuales con formatos accesibles, facilitan su distribución y análisis, lo que contribuye a la producción de nuevo conocimiento. Siguiendo con la misma línea que destaca el Open Data sobre el acceso a la información se encuentra la posibilidad de consultar patentes para el desarrollo de proyectos en el caso de México, la manera de examinar estas patentes es a través del Instituto Mexicano de la Propiedad Industrial (IMPI) complemento del Sistema de Información de la Gaceta de la Propiedad Industrial (SIGA). En la cual se puede acceder y consultar la información sobre las 1,555 solicitudes de patentes que tiene México hasta 2018.<sup>14</sup>

Por otro lado, la creación y difusión de los portales de transparencia que tiene el gobierno federal no es la única fuente con datos abiertos existente, como se ha mencionado, el Open Data, al tratarse de bloques de información que forman parte del Big Data muchos de los datos que se encuentran en internet en su mayoría pueden ser consultados sin restricciones, así como los sitios web dedicados a la divulgación de información científica dirigidos principalmente por instituciones académicas, esto, con la intención de que los datos puedan beneficiar y satisfacer las dudas del usuario.

Uno de los ejemplos más representativos que tiene el Open Data son los buscadores web, ya que en ellos circula prácticamente toda la información que es generada diariamente, lo que ha hecho que se convierta en una fuente potencial de recolección de información, tanto así que uno de los navegadores más conocidos ha logrado trascender como una expresión (googlear). Sin embargo, Google no es el único que mantiene un modelo de apertura de datos. Tal vez el ejemplo con mayor afinidad a nivel de usuario es el caso de las páginas de distribución de código abierto, en donde los propios usuarios de la plataforma suben archivos sobre proyectos de programación con la intención de que pueden ser editados, desarrollados, incluso complementados por la comunidad. Teniendo así a páginas como Github, Fosshub, Sourceforge como los más populares, además de ello la

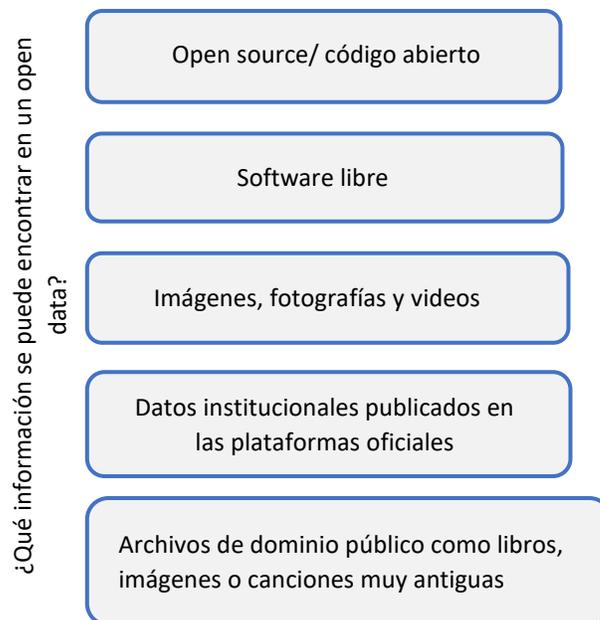
---

<sup>14</sup> Cifra tomada del informe 2018 del IMPI en <https://bit.ly/2sLIBUd>

existencia de las plataformas libres, es decir, aquellos sitios que ofrecen sus servicios o contenido sin ningún costo.

Dicho de otro modo, el Open Data ha sido una herramienta con gran utilidad para la sociología, debido a que gracias a las políticas de transparencia gubernamentales y los métodos de descarga libres de pagos, es posible obtener bases de datos, imágenes, mapas, documentos o casi cualquier tipo de información que se encuentre en internet, así como también software que puede ser utilizado para facilitar el trabajo del investigador. Sin embargo, no tan sólo la posibilidad de adquirir herramientas sin costo alguno da la ventaja al sociólogo sobre el valor de los datos, realmente la ventaja adquirida del investigador con los datos abiertos son por la relevancia de la información que puede ser encontrada en todo un océano digital, es decir, en la mayoría de las ocasiones cuando se inicia un proyecto de investigación, se suele buscar material relacionado con el tema, por lo que tener la posibilidad de consultar información gubernamental y contrastarla con la de otros países o revisar las producciones académicas en otras instituciones puede ser un factor para el desarrollo de la investigación.

**Figura 7. tipo de información encontrada en Open Data**



Fuente: Elaboración propia

### **2.3.2 Social Data**

A diferencia del Open Data que permite la consulta de información mediante los modelos de transparencia gubernamental, la publicación de producciones académicas y el acceso libre a casi cualquier información mediante internet, el Social Data se encuentra enfocado principalmente en la recolección y al análisis de los datos provenientes de las redes sociales, los cuales son usados para establecer patrones de comportamiento o preferencias mediante la información de los usuarios. Esto en principio puede significar una mayor complejidad en el tratamiento de los datos, pero también una mayor recompensa debido a que los repositorios de datos iniciales pueden arrojar mejores resultados durante análisis, aumentando así el valor de los datos para el usuario/investigador. Por lo tanto, dentro de la jerarquización que aquí se presenta, el Social Data es considerado como el segundo nivel de profundidad que tiene Big Data.

Algunos de los usos que tiene el Social Data han sido concebidos en la mercadotecnia, la publicidad y la política, como parte de ese nicho encargado de la construcción de campañas mediáticas de alto impacto, presentes en la industria de la moda y las campañas electorales. "... las redes sociales digitales constituyen un complejo entramado tecnológico, económico y político que les permite transformar la narrativa personal, los gustos y los estados de ánimo en datos, los cuales son su motor económico y base de su modelo de negocio. A pesar de la brecha digital que deja fuera a 51.8% de los habitantes del mundo, la vida personal y laboral, así como la vida política, transcurre en buena medida en las redes sociales digitales" (Meneses, 2018: 417). Es decir, los datos emanados de estas redes permiten que se pueda perfilar a los usuarios con el fin de establecer públicos objetivos. Sin embargo, la principal característica que tiene el Social Data es que se basa en la medición de aceptación, propagación y reacción de los usuarios principalmente dentro de las redes social sobre los productos, temas o problemáticas sociales. por lo que son capaces de personalizar el contenido, trayendo consigo la masificación de un mensaje, a través de elementos particulares que ayudan a obtener una mayor reacción sobre las personas. Por ello, este nivel se encuentra estrechamente ligado

con el análisis de sentimientos y el análisis de redes sociales, debido a que su uso se deriva del diseño y monitoreo de campañas a través de la social media<sup>15</sup> para conocer el comportamiento de un tema en particular.

Ahora bien, "... los social media, caracterizados por su rápida expansión y por la variedad de interacciones de contenidos que en ellos circulan, se han incorporado al estudio del big data al convertirse en fuente de datos útiles para investigadores, entidades y empresas" (Martínez y Lara, 2015: 575). Visto desde una perspectiva sociológica, el análisis de estos datos puede ser de gran ayuda para el monitoreo de eventos coyunturales, la evolución de movimientos sociales e inclusive el seguimiento de la opinión pública sobre problemáticas sociales específicas, ya que "... se pasó de las redes sociales, que hemos visto hasta ahora, al emergente campo del análisis de los big data y a la interacción entre ambas perspectivas que se refuerzan mutuamente" (Elías, 2016: 1142). Ejemplos de ello son las campañas electorales de Barack Obama en 2008 y 2012, las elecciones presidenciales de EUA en 2016 y las campañas electorales de México en 2018, donde el diseño y la construcción de las campañas fueron basadas en una infraestructura de redes sociales con el fin de atraer la mayor cantidad de la población electoral.

Por si fuera poco, con la llegada del Big Data y el análisis de todo el contenido refugiado en las plataformas digitales ocurrió "... uno de los fenómenos más interesantes de los últimos años, de la mano de la informática, es el regreso de las matemáticas a disciplinas que se habían separado de ellas -y estaban en el ámbito de la antigua filosofía y letras- como la política, la sociología o el propio periodismo" (Elías, 2016: 1142). Es decir, el intercambio de datos de todo tipo sumado a la incorporación de las matemáticas y la informática en las interpretaciones de las ciencias sociales dio como resultado el surgimiento de nuevas corrientes que contribuyen a la interpretación de nuevos paradigmas en la era digital.

Por lo que es de esperar que surjan nuevos conceptos que ayuden a definir la manera en que interactúa la sociedad a través de la digitalización y con ello el

---

<sup>15</sup> Se refiere al conjunto de redes sociales existentes en la actualidad.

funcionamiento del Social Data y su contribución a la sociología. Iniciando con el concepto de *ciberrealidad*<sup>16</sup> que se define como “aquella realidad real que está condicionada por lo que sucede en el mundo virtual.” (Elías, 2016: 1134). Dicho de otro modo, la existencia de una vida en el mundo virtual que convive de forma paralela con la ya existente en el mundo real permite que se creen diversas interpretaciones sobre su personalidad, haciendo que el individuo inmerso en el mundo virtual deba utilizar máscaras para no verse vulnerables ante los otros, construyendo así fachadas que se adecuen a la dependiendo a escenarios presentes en su contexto, al igual que lo menciona Goffman en *la presentación de la persona en la vida cotidiana*. Ejemplo de ello, es la transformación efectuada en la esfera política con la incorporación de las redes sociales que son usadas principalmente para construir la imagen del político para atraer adeptos que coincidan con sus ideales en campaña.

Teniendo en cuenta este paralelismo en la realidad del individuo “...los blogs, los correos electrónicos y el Twitter promueven una conexión instantánea, sin fronteras, que permite que unos ideólogos de mentalidad parecida se coordinen con personas a las que no conocen a través de un portátil a altas horas de la madrugada o desde un iPhone mientras llevan los niños al parque” (Elías, 2016: 1141). En otras palabras, formando afinidades con otros individuos a través de los medios digitales, con la posibilidad de que puedan ser interpretadas por la sociología. Cabe mencionar que “...el mundo virtual no tiene por qué funcionar como el real, pese a que el real está influido por el virtual” (Elías, 2016: 1134). Como se puede notar actualmente con las tendencias producidas en las redes sociales.

No obstante, “... el individuo como ser social experimenta la necesidad de contar y narrar sus vivencias y ahora esa necesidad se traslada a internet gracias a las posibilidades que brindan los medios sociales” (Martínez y Lara, 2015: 576). Es decir, el hecho de tener alguna interacción en la World Wide Web hace que todos

---

<sup>16</sup> El concepto es establecido por Carlos Elías (2016) para describir la evolución de la comunicación política en la era digital, mencionando la utilización de las redes sociales para la promoción de los candidatos, así como para la atracción de electores.

los datos producimos a través de la red, sean vinculados a un rastro digital dejado por nosotros al navegar, hecho que ha resultado en la formación de una identidad virtual, algo parecido a lo que menciona Carlos Elías (2016) con el *second life* refiriéndose principalmente a la imagen construida por los políticos a través de las redes sociales. Sin embargo, independientemente del uso político que tiene, el individuo una vez que es capaz de navegar, comienza a desarrollar su propia ciberpersonalidad, por lo que "... la datificación en social media se relaciona entre otros con los mensajes y contenidos publicados y/o compartidos, los estados de ánimo o gustos expresados (a través de los botones "me gusta" o "favorito"), la descripción de los perfiles los sentimientos, experiencias profesionales y aspectos aparentemente imperceptibles recogidos en forma de metadatos" (Mayer-Schönberger; Cukier, 2013, p. 166; Martínez y Lara, 2015: 577). Haciendo que la información que anteriormente no era posible medir se vuelva medible a través de aspectos sociométricos.

"Los medios sociales han tenido un espectacular incremento en los últimos años y han marcado una revolución en los hábitos y posibilidades de comunicación, intercambio y publicación de contenidos" (Martínez y Lara, 2015: 576). Trayendo consigo la necesidad de interpretar el comportamiento de los usuarios en las redes, por lo que "...el análisis de los datos generados en los medios sociales se presenta como un nuevo campo de estudio en el que se hace necesaria la aplicación de nuevos métodos y tecnologías" (Martínez y Lara, 2015: 578). Y a pesar de que la sociometría no es un método de reciente creación, la incorporación para develar los misterios que tiene la interacción de las personas en internet desata una nueva perspectiva. Por ello, dentro del análisis de las redes es necesaria la sociometría<sup>17</sup>, ya que dicho método se basa en el estudio del comportamiento de individuos dentro de sectores sociales o bien grupos, anexando también la teoría de grafos<sup>18</sup> para

---

<sup>17</sup> La sociometría es un método de investigación desarrollado por Jacob Levy Moreno en 1932, dicho método involucra la medición de factores psicosociales para analizar las estructuras sociales establecidas de forma cuantitativa. Para más información sobre el tema puede consultar en <https://bit.ly/2Qp0Fg4> o también <https://bit.ly/37S1ALS>

<sup>18</sup> La teoría de grafos es una rama de las matemáticas aplicada que tiene sus fundamentos en el álgebra, geometría analítica, aritmética y estadística, y es recurrentemente utilizada en las ciencias de la computación

ayudar a formar modelos gráficos (sociogramas) que se adecuen a la explicación. Es decir, la aplicación de análisis exploratorios que muestren las relaciones interpersonales de los individuos al navegar en internet.

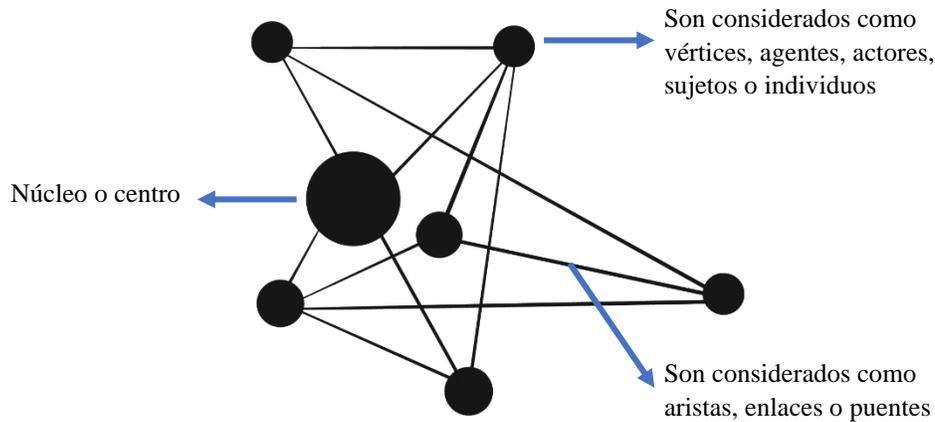
En conjunto la sociometría, el análisis de redes sociales y la teoría de grafos generan los elementos adecuados que permiten descifrar e incluso interpretar un poco más del comportamiento de las personas en internet, de allí viene la importancia de analizar el Big Data y el Social Data desde una perspectiva sociológica. Ahora bien, antes de continuar se deben conocer los elementos que forman una red, esto debido a que la interpretación que el sociólogo pueda hacer se encuentra basada en los modelos emanados de los metadatos extraídos de ellas.

Una red, de acuerdo con la teoría de grafos, se encuentra compuesta por un grupo de puntos también llamados vértices o nodos y un grupo de líneas llamadas aristas o enlaces, que sirven para conocer la interacción que tiene entre sus componentes. Es decir, una red es un mapeo de información que sirve para conocer el grado de relación que tienen las personas, los grupos sociales. en otras palabras, es un mapeo de la estructura social digital de las personas y que se encuentra ligada a su identidad virtual. Dentro de esas estructuras se encuentran los nodos o vértices que indican la presencia de una persona, individuo, sujeto o actor dentro de la red y usualmente se encuentra representada por un número o un punto en la unión final de las líneas. Por otra parte, también se encuentran los enlaces, que indican la existencia de una interacción entre dos nodos o más dentro de la misma red, dicho de otra manera, es el vínculo entre dos o más personas pueden tener y se encuentra representada por líneas definidas entre los puntos relacionados. A su vez estas líneas pueden representar la emisión o recepción de un mensaje (véase figura 8.)

---

para realizar análisis sobre las propiedades de los grafos. Para más información sobre el tema puede consultar en <https://bit.ly/2uA9fQv>

**Figura 8. Elementos de una red**

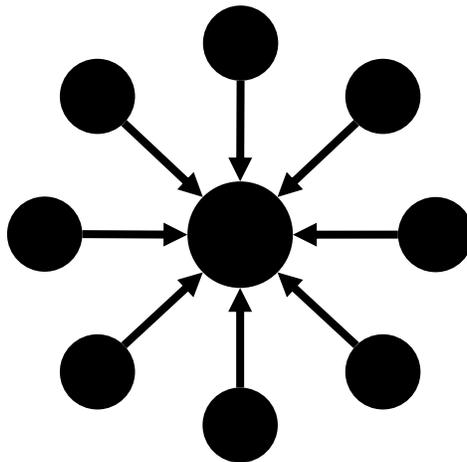


Fuente: Elaboración propia

Aunado a lo anterior, existen diversos tipos de relaciones que facilitan la manera de interpretar las interacciones interpersonales de los individuos en internet, dichas relaciones pueden ser establecidas como *simétricas* (dirigidas) o *asimétricas* (no dirigidas), esto quiere decir que la estructura social modifica el grado de relación e incluso el mensaje entre el emisor y receptor.

Una *relación simétrica o dirigida* dentro de una red son relaciones de una sola vía, por lo que el mensaje entre el emisor y el receptor es en un solo sentido. Por ejemplo, un creador de contenido en YouTube que se graba así mismo no es capaz de recibir una respuesta sobre el video que grabó hasta después de haberlo subido a la plataforma; otro ejemplo es el de un profesor explicando un tema a sus alumnos, es decir, el profesor se encuentra interactuando con sus alumnos sin que haya de vuelta una respuesta, por lo que el mensaje entre emisor – receptor se vuelve de una sola vía. Cabe mencionar que este tipo de relación es una relación binaria, ya que son relaciones sin peso y se encuentran representadas gráficamente por una flecha. (véase figura 9.)

**Figura 9. Relación simétrica**



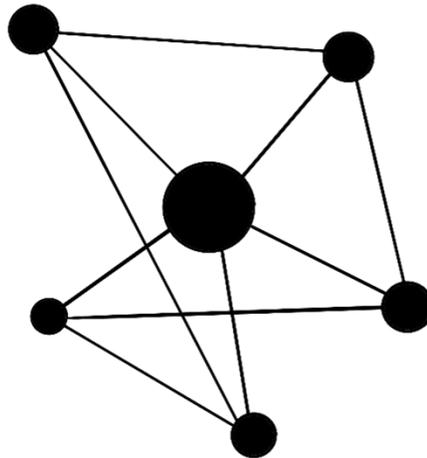
Fuente: Elaboración propia

En el caso de las *relaciones asimétricas o no dirigidas* son aquellas que indican la existencia de una relación/interacción entre dos o más nodos, es decir, son relaciones que tienen peso por lo que tienen valores asociados que indica la fuerza o la frecuencia del vínculo expresado. Gráficamente estas relaciones son expresadas con líneas con mayor o menor grosos para expresar el grado de relación existente entre los nodos. Por ejemplo, el grado de relación que se tiene un usuario con su lista de amigos en Facebook, es decir, una vez formado el grafo se muestra que los vínculos de mayor grado o importancia son los usuarios con los que tiene contacto en la vida real y los de menor importancia son los que no tiene contacto directo o solo tiene contacto a través de la red.<sup>19</sup> Otro ejemplo es una persona con la solicitud de libros en las bibliotecas, en donde se puede notar que tipo de libros suele solicitar, genero, autor, el tiempo de préstamo e inclusive las bibliotecas de donde son prestados. (véase figura 10.)

---

<sup>19</sup> cabe aclarar que en algunas ocasiones no tener un contacto directo no significa que los sujetos no puedan formar afinidades que aumenten el grado de importancia de interacción.

**Figura 10. Redes asimétricas**



Fuente: Elaboración propia

Así como existen distintos tipos de relación dependiendo si los enlaces tienen peso o no, también existen distintos tipos de redes que permiten al investigador establecer la correlación más adecuada para hacer su interpretación, como se muestra en la tabla 1.

<b>Tabla 1. Tipos de redes</b>	
<b>1. Redes complejas</b>	Son aquellas que muestran todo tipo de vínculos o interacciones entre los nodos establecidos en la red. Por ejemplo, internet es una clara representación de una red compleja, debido a la interconexión de los servidores distribuidos por todo el mundo; Algo más cercano a la sociología que represente una red compleja son las ciudades.
<b>2. Redes parciales</b>	Son aquellas que muestran solo una parte de la red, es decir muestra las relaciones, los vínculos o las interacciones más significativas dentro del grafo, así como también las ramificaciones formadas dentro de la red. Ejemplo de ello, son las variaciones existentes en el idioma español, tanto regional como de manera internacional.

<b>3. Redes temporales</b>	<p>Son aquellas que analizan periodos específicos de tiempo que son determinados por el propio investigador. Un ejemplo de esta son las líneas de tiempo que establecen la relación con cada uno de personajes, otro ejemplo es el análisis del periodo de aparición de una obra literaria.</p>
<b>4. Redes multimodales</b>	<p>Son aquellas que analizan un tipo de relación que sea de interés para el investigador, en estas redes sólo se incluye un tipo de vértice donde se establece la relación entre personas, objetos, temas, pero no entre ellos.</p>
<b>5. Redes de afiliación</b>	<p>Son las que establecen algún tipo de afinidad entre los sujetos pertenecientes a la red. Esta última, usada comúnmente en la mercadotecnia para relacionar la conducta de las personas con objetos específicos. Por ejemplo, el caso de una campaña publicitaria sobre el nuevo smartphone de la compañía de la manzana mordida.</p>
<b>6. Redes múltiples</b>	<p>Son aquellas que poseen distintos tipos de redes dentro de la red principal. Ejemplo de ello, son las conexiones que tiene una persona a lo largo de su vida, en donde cada uno de los individuos con los que se encuentra conectado tiene una red propia que puede o no interconectarse con él.</p>
<b>7. Red egocéntrica</b>	<p>Son aquellas redes que son formadas sólo por aquellas personas conectadas a la persona que nos interesa.</p>

Fuente: Elaboración propia

Como lo habrán notado, cada una de las redes que aquí se describen brinda la oportunidad al sociólogo de interpretar la interacción entre los nodos de la mejor manera posible. Dicho lo anterior, cada parte de la red una vez que se comienza a descifrar es concebida como un nuevo elemento, volviéndose un miembro, es decir, de los puntos/nodos son transformados a centros, puentes, mediadores o aislantes en función del tipo de interacción que tengan entre sí, esto, asociado a la teoría de

rol; mientras que los enlaces se mantienen delimitando el grado de relación que tienen los miembros de la red, como se muestra en la tabla 2.

<b>Tabla 2. Miembros de la red</b>	
<b>1. Centro</b>	Como su nombre lo indica, es el nodo central que nos interesa analizar dentro de la red, cabe mencionar que puede haber más de un nodo principal y pueden ser notados por la popularidad, es decir, el número de enlaces existentes entre cada uno de ellos con los demás miembros de la red.
<b>2. Puentes</b>	Son los miembros que tienen múltiples ramificaciones en la red, estos miembros suelen ser el vínculo con otros grupos sociales o clústeres dentro de la misma. Es decir, son aquellos nodos que poseen más de un enlace.
<b>3. Árbitros o mediadores</b>	Son los miembros que poseen vínculos con redes externas a la red analizada.
<b>4. Aislantes o ermitaños</b>	Son aquellos individuos que pertenecen a una red, pero que se encuentran aislados del resto, estos miembros suelen poseer pocos o nulos vínculos con los demás miembros de la red.

Fuente: Elaboración propia

Cabe destacar que la estructura social, al encontrarse en un constante cambio los patrones de comportamiento e incluso las relaciones que se muestran en los sociogramas, tienden a modificarse debido a la rápida expansión en las redes sociales “... los datos por sí mismos pueden no aportar nada o incluso generar ruido, especialmente si se encuentran en volúmenes elevados, por lo que su procesado se hace indispensable” (Martínez y Lara, 2015: 576). Es por ello, que el análisis de redes en la era de la digitalización es esencial para la decodificación e interpretación del comportamiento social de los individuos en internet.

Ahora bien, dejando un poco de lado la parte técnica de los elementos que conforman una red. “... el valor del análisis de la social media data se observa no sólo en la inversión que se está asignando a su investigación, sino también en su utilidad para, entre otras cuestiones, comprender el pensamiento y actuación de las personas, mejorar los procesos que llevan a la toma de decisiones y dirigir de forma eficaz productos y servicios” (Hobbs, 2014; Martínez y Lara, 2015: 576). Visto desde una perspectiva sociológica, da como resultado una mejor interpretación sobre los motivos de actuación de las personas dentro de un entorno digital y que suele ser estudiado no tan sólo con aspectos cuantitativos, sino también recurriendo a la parte cualitativa, que “...desde la perspectiva de la investigación se trata de adaptar y aprovechar metodologías, algoritmos y funciones existentes para el análisis” (Martínez y Lara, 2015: 578). Aplicando en conjunto lo aprendido en el análisis estadístico con elementos subjetivos como los sentimientos.

En otras palabras, para poder conocer el comportamiento social de una identidad virtual desde una perspectiva sociológica “se requiere el uso combinado de grafos para detectar relaciones y perfiles similares y contenidos afines. Esto permite el cálculo de relaciones de la red social a partir de métricas que incluyen entre otros el análisis de authorities y de hubs que establecen la importancia de los nodos y su relevancia respecto al resto de nodos” (Martínez y Lara, 2015: 578). Lo que se puede traducir en una serie de dimensiones que permiten examinar cómo se encuentra constituida la red. (véase tabla 3.)

<b>Tabla 3. Dimensiones de una red</b>	
<b>1. Contenido transaccional</b>	Refiriéndose al tipo de intercambio efectuado en la red y que permitirá identificar el tipo de relación existente entre los miembros que puede ser de afecto, influencia, información, recursos, bienes o servicios.
<b>2. Naturaleza de los nexos</b>	Se refiere a la intensidad o la fuerza de los vínculos existentes entre los miembros de la red. Es decir, el grado de afinidad existente entre los nodos de la red

<b>3. Reciprocidad</b>	Se refiere a la percepción de las relaciones entre los miembros de la red, en otras palabras, es el grado de importancia del enlace.
<b>4. Multiplicidad</b>	Es el grado de relación que tienen los miembros de la red, refiriéndose al grado de cohesión o afinidad existente en la red.
<b>5. Tamaño</b>	Se refiere al número de personas en la red. <sup>20</sup>
<b>6. Densidad:</b>	Se refiere al número de nexos/enlaces reales proporcional al número total de nexos posibles dentro de la red.
<b>7. Agrupación</b>	Se refiere al número de regiones densas que son mostradas en la red o también llamados clústeres/vecindarios
<b>8. Centralidad</b>	Se refiere al grado jerárquico de los miembros, además de la restricción en la comunicación en la red.
<b>9. Estabilidad</b>	Se refiere al grado con que el patrón de la red interactúa con el tiempo.

Fuente: Elaboración propia

### **2.3.3 Geo Data**

Hoy en día, en una época en donde la más mínima interacción con los aparatos electrónicos genera mediciones e información, tener la oportunidad de acceder a todo tipo de datos de forma libre y gratuita puede significar un gran avance para quien puede aprovechar la información. Por ello, continuando con la jerarquización que aquí se presenta, el Geo Data se toma en consideración como el tercer nivel

---

<sup>20</sup> Nota: para obtener resultados óptimos en el análisis de la red se sugiere la conformación de mínimo 50 miembros en la red, debido a que una red con menos miembros puede ser confusa la interpretación, mientras que entre mayor número de miembros el análisis puede arrojar mejores resultados.

dentro de Big Data, debido a que recopila los datos geolocalizados emitidos por dispositivos electrónicos con una conexión de red y que pueden ser utilizados con múltiples propósitos.

Al igual que sucede con el Social Data, el Geo Data toma principalmente los datos que se encuentran dentro de las redes sociales para después crear marcas de movimiento. Sin embargo, no sólo esos datos son los que alimentan a los Sistemas de Información Geográfica (SIG), sino que “Big (Geo) Data cubre ámbitos muy distintos: actividad de los teléfonos móviles, registros de operaciones con tarjetas de crédito, datos recogidos en tiempo real con GPS, redes sociales, registros de consumo de agua y electricidad, imágenes grabadas con cámaras de video, tarjetas de transporte público o de uso de sistemas públicos de bicicletas, etc.” (Gutiérrez-Puebla et al, 2016: 6). Inclusive en el momento que nos encontramos caminando nuestros dispositivos inteligentes se encuentran emitiendo señales buscando una conexión y que terminan marcando nuestra ubicación geográfica. Dado que “...una buena parte del Big Data contiene datos geolocalizados, éstos pueden ser tratados y analizados con Sistemas de Información Geográfica y técnicas de estadística espacial, además de con herramientas de la estadística clásica” (Gutiérrez-Puebla et al, 2016: 5). Por lo que son usados principalmente para analizar distintos aspectos de la sociedad, como lo es la movilidad en las ciudades.

Sí hablamos de la manera en que puede ser utilizada la metadata geoespacial, podemos encontrar que, en su mayoría, las aplicaciones e incluso los dispositivos requieren recopilar algún tipo de dato que posteriormente formará parte de un Sistema de Información Geográfica (SIG). “Esta necesidad de datos es muy elevada en el caso de los SIG, que pueden convertirse en una herramienta versátil y llena de posibilidades sobre unos datos geoespaciales de calidad, o bien, ser de escasa o nula utilidad en caso de que exista una carencia de datos” (Olaya,2009: 3). Siendo esta información comúnmente utilizada como una forma de analizar el movimiento geoespacial de las poblaciones a través de los geodatos.

Cabe mencionar que actualmente a través de aplicaciones de mapas de diversas compañías los usuarios pueden acceder a los registros sobre cada punto donde han estado y revisar todo el tracking. (véase figura 11.) Ahora bien, ¿Qué es un geodato? un geodato es considerado como “cualquier elemento u objeto tangible que existe en la realidad y que posee una ubicación o posición geográfica, la cual es normalmente almacenada como coordenadas geográficas con topología, aunque también puede ser implícita” (Falla Gamboa, 2012: 4-5). Es decir, que es una marca dentro de un mapa virtualizado, la cual puede ser utilizada como parte del análisis de una investigación. Sin embargo, “...los datos geoespaciales plantean su propia problemática, y al igual que en el caso del software, existen diversos enfoques en lo que respecta a su distribución y uso, no sólo desde el punto de vista de las restricciones y licencias a aplicar, sino también desde el punto de vista técnico. Ambos están íntimamente ligados, y existe una relación directa con los modelos de distribución del software” (Olaya, 2009: 3). Es decir, que al igual que se presentan las diversas problemáticas con el Open Data con la restricción de información, y en el Social Data con la dificultad en la manipulación de la meta data extraída de las redes sociales, el Geo Data presenta en gran medida la manipulación de datos concentrados en repositorios de información centralizada que llegan a requerir software especializado, limitando su posible análisis. No obstante, el no poder acceder al software especializado no es impedimento para analizar los datos en esos repositorios, por lo que nos permite recurrir a aplicaciones de open source que tienen con las mismas capacidades que el de distribución original y que pueden funcionar para el sociólogo cuando se trata de analizar información.

En cuanto al análisis de información geolocalizada “...las fuentes de datos disponibles permiten a los investigadores poner fin a su dependencia de las estadísticas oficiales, en campos tan diversos como la demografía, la actividad económica, la movilidad y los flujos, y cualquier otro de los aspectos urbanos” (Shelton et al, 2015; Gutiérrez-Puebla et al, 2016; 5). Es decir que se vuelve una línea de investigación prometedora para la sociología. Por otro lado, “...el uso de Big Data también permite medir la concentración de personas en protestas sociales” (Conover et al., 2013; Gutiérrez-Puebla et al, 2016: 8). A través de la meta data de

las redes sociales, que al igual que se ha visto con el Social Data, el contenido subido a las redes sociales puede ser analizado para determinar la progresión y el crecimiento de estos movimientos.

**Figura 11. Registro de actividad en Google**



Fuente: Retomado de Google en <https://myaccount.google.com/activitycontrols>

Nota: Por motivos de privacidad no se pueden mostrar las ubicaciones del usuario que prestó la cuenta. No obstante, se presenta esta imagen como un comprobante sobre el registro continuo que se hace sobre nuestra ubicación y el cual puede ser consultado a través de la configuración de privacidad de Google.

Considerando que gran parte de la información encontrada en Big Data se encuentra generada por las redes sociales y que a su vez estos datos contienen información geolocalizada, son creados los algoritmos de comportamiento como se relata en el apartado anterior, en el caso de Twitter, una red social que suele ser utilizada a manera de blog y en el cual se puede compartir la actividad que realiza cada usuario es posible notar que "...cada tweet deja una "huella" digital del lugar y el momento en que fue emitido" (Gutiérrez-Puebla et al, 2016: 7). Por lo que se

pueden crear estimaciones sobre los sitios que ha visitado cada usuario. Teniendo en cuenta lo anterior, según Gutiérrez Puebla (2016) tan sólo alrededor del 3% en 2015 y el 1% en 2016 de los datos generados en Twitter son datos de los que se conocen las coordenadas del emisor.

Ahora bien, ejemplos de ello además de Twitter se encuentra Google con los registros de actividad diaria y en el cual se pueden consultar nuestras ubicaciones e incluso nuestros recorridos, asimismo también podemos encontrar las aplicaciones de transporte como Uber, Cabify, DiDi o Beat, en donde el funcionamiento de dichas aplicaciones prácticamente se encuentra basado en el envío y recepción de coordenadas de una persona para que pueda ser transportado, esto mismo sucede cuando llegamos a utilizar los servicios de mapas como el de Waze o de Google para consultar una dirección, el tráfico o compartir, nuestra propia ubicación. Es decir, se forman perfiles espaciotemporales que están ligados a coordenadas con las ubicaciones registradas de los usuarios, mismos que pueden ser usados como una forma de monitoreo, que en la medida de lo posible pueden ayudar a los investigadores para dar un seguimiento de los fenómenos sociales a través de la selección de información como los hashtags en el caso de Twitter, o bien, con los puntos de interés en Foursquare, así como el tráfico existente en la zona en alguna aplicación de mapas. De esta forma se pueda analizar la evolución de un tema que puede ser coyuntural, la intensidad de los mensajes publicados, además del nivel de movilización dentro de cualquier red social.

Como se pudo apreciar en este segundo capítulo, el Big Data logró adquirir una gran importancia desde el momento que las empresas comenzaron a buscar soluciones rápidas y efectivas que se adecuaran al mercado actual, logrando que otros sectores (entre ellos el publicitario, político, salud y el académico) se viesen interesados en estudiar el comportamiento social a través de los grandes datos con fines particulares. Asimismo, se presentan los distintos niveles de profundidad perceptibles en Big Data (Open, Social y Geo Data) cada uno con características particulares que aumentan su dificultad para ser analizados.

### Capítulo 3. El Big Data desde una perspectiva sociológica

Por lo que en este último capítulo resta por conocer la integración y realización de un análisis de Big Data enfocado hacia una perspectiva sociológica. Cabe mencionar que dicha integración no se resume a que el investigador deba o no tener un vasto conocimiento en programación e informática, sino que depende de la capacidad del sociólogo para poder procesar la información con las herramientas que nos provee la sociología, sin dejar de lado en la medida de lo posible, el conocimiento de lenguajes de programación como Python<sup>21</sup> que facilitará el trabajo al momento de incorporar el uso del Big Data como parte de su investigación sin importar si sea de corte cualitativo o cuantitativo.

Si bien, para un científico de datos o un informático resulta mucho menos complicado entender el Big Data, y asimismo asimilar la forma de realizar un análisis debido tal vez a su formación especializada, para un sociólogo “... el análisis de grandes volúmenes de datos puede tener un valor para el conocimiento de lo social, y para lo cual es imprescindible que los científicos sociales realicen sus propios experimentos, no sin antes reconocer las condiciones sociotécnicas en que los datos fueron generados” (Meneses, 2018: 430). En otras palabras, es necesario tomar en cuenta que para analizar el flujo constante de datos que provee Big Data, se debe tener la claridad suficiente para abordar el fenómeno, tomando en cuenta que “...el análisis con macrodatos mediante técnicas analíticas, como el aprendizaje automatizado (Machine Learning) o la minería de datos (Data Mining), permite encontrar patrones con los que se pueden elaborar modelos útiles para el pleno entendimiento de lo social” (Meneses, 2018: 423). Es decir, hoy en día casi cualquier elemento generado en la era digital puede ser convertido en una fuente potencial de datos (redes sociales, aplicaciones, navegadores web, etc.) y del que los sociólogos al hacer uso de estas nuevas fuentes podrían sacar provecho para medir el comportamiento social, esto último también toma en consideración no solo

---

<sup>21</sup> Aclarando este punto, se recomienda Python como parte de los conocimientos básicos para realizar análisis de Big Data debido a que en su mayoría los programas indexados a la parte estadística y de análisis de datos permiten en cierta medida este lenguaje, además que algunos programas como SPSS se encuentran basados en Python.

aquellos datos que son de corte cuantitativo, sino también de corte cualitativo como los discursos, narrativas e imágenes.

Ahora bien, pese a que la brecha digital se ha escudriñado con el análisis y el control de los datos, resulta para los científicos sociales en el abordaje inevitable de temas que no suelen ser tratados usualmente en la sociología, por lo que algunos de los análisis de Big Data han sido también vinculados al plano de las ciencias sociales tomando en cuenta que "... los métodos digitales son de gran ayuda para el investigador y en ocasiones resultan útiles para seleccionar el corpus de un estudio en cualquier plataforma digital con criterios menos subjetivos" (Meneses, 2018: 433). Siempre y cuando se procure que la selección de temas que sean de interés para la sociología, es decir, que afecten de alguna forma las estructuras sociales, o bien, que sean temas coyunturales que en su mayoría se desprenden de la opinión pública.

Con el aprendizaje de lenguajes de programación, ya no sólo por parte de los matemáticos y de los informáticos, además de la creciente explotación de los grandes repositorios de información "... Dominique Boullier (2016), investigador del Laboratorio de Ciencias Políticas de la Universidad de París, ha llegado a comparar la ola de Big Data con otros momentos de la sociología y su dependencia de las encuestas para explicar lo social" (Meneses, 2018: 424). Refiriéndose a que la búsqueda de la sociología por contar con pruebas suficientes del cambio de las estructuras sociales ha dejado como resultado que los nuevos científicos sociales recurran a incorporar nuevas formas de recolectar datos que generen información relevante para sus estudios y así poder analizar patrones de comportamiento social que no son notables a simple vista. Con esto último no me refiero a que la llegada del Big Data a la sociología sea el principio de una etapa efímera dentro de la disciplina o que se encuentre estancada, sino que la modificación de estrategias y formas de trabajo del investigador para configurar el entramado teórico metodológico de su estudio ha significado un avance en la incorporación del análisis de grandes datos en la sociología.

Se podría decir que a partir de los grandes datos presenciamos un momento en el cual se abren posibilidades en la imaginación de los sociólogos para la observación de otros fenómenos sociales, ya que amplía la posibilidad de observar y experimentar con nuevos campos de estudio dentro de la disciplina. Es decir, "... la relevancia de este tipo de ensamblaje para las ciencias sociales no estaría dada por atributos relacionados con el tamaño y especificidades técnicas, sino por la forma en que se generan, en las relaciones con otros datos, por las conexiones de un individuo, de un individuo con otros, de grupos sociales, o bien, de la propia información" (Meneses, 2018: 425). De esta forma, los límites impuestos tradicionalmente en la formación de los científicos sociales se van desvaneciendo con el abordaje de problemas complejos desde una perspectiva multidisciplinaria. Cabe aclarar que el hecho de que el investigador deba tener una perspectiva multidisciplinaria no se refiere a que, con cada proyecto planteado u objeto de estudio nuevo, deba formar equipo con personal especializado, aunque pueda ser de gran ayuda en casos específicos, sino se refiere a que el propio investigador/científico social/sociólogo, debe contar con técnicas o herramientas independientes a su formación que, en la medida de lo posible, le ayuden a explicar de una mejor forma la realidad social.

Por otra parte, "... la falta de entrenamiento de los científicos sociales puede traducirse en una claudicación ante las ciencias computacionales en favor de un empirismo que desplace las categorías de pensamiento de las ciencias sociales" (Meneses, 2018: 429), lo que hace que puedan surgir otros obstáculos o dificultades con el tratamiento de los grandes datos como sesgos en la información o la neutralidad del investigador respecto a los paquetes de datos analizados. Ya que en ocasiones medir patrones no significa poder explicar el comportamiento de los actores, esta advertencia resulta útil para aquellos nuevos investigadores que pretenden incursionar en el análisis de metadatos, o bien, que están en plena formación, refiriéndose a que el análisis de datos por sí sólo no exime a los investigadores de fundamentar sus interpretaciones de las estructuras sociales cambiantes o los motivos del constante cambio social, por lo que la teoría resulta indispensable para dar una explicación de los fenómenos y que en ocasiones, "...

los algoritmos con que se escarban las huellas más profundas de los datos no siempre son neutrales” (Meneses, 2018: 425), provocando que la objetividad y la veracidad de la información recabada pueda no ser concluyente o congruente con los resultados.

### **3.1 Ecosistemas digitales para utilización del Big Data en la sociología**

La sociedad se transformó, y con ello también se han ido modificando las estructuras sociales haciendo que la labor del sociólogo recaiga en buscar nuevas formas de analizar la sociedad, es decir, como lo mencionaba Castells en su momento, la sociedad ha adoptado un estilo de vida digital, haciendo que las prácticas sociales cambien, pasando de los métodos tradicionales de comunicación y consulta de información a una sobreproducción de datos proveniente del desarrollo tecnológico de las últimas décadas.

Dicha datificación se vuelve un punto itinerante para el investigador según este paradigma, debido a que los sociólogos deben modificar la perspectiva con la que perciben al mundo, ya no tan sólo de manera física sino incorporando también la identidad virtual de los individuos para estructurar de mejor forma sus interpretaciones. Por ello, se establecen biomas digitales que le permitan al sociólogo trabajar con las grandes cantidades de datos que son producidos diariamente. Con esto en cuenta no se hace referencia a que los sociólogos también deban ser ingenieros en computación o científicos de datos, a menos que así lo deseen, sino que dentro de su formación consideren adquirir conocimientos básicos en dichas disciplinas para poder llevar a cabo de mejor forma el tratamiento de los datos con *suites de cómputo*<sup>22</sup> distintas a SPSS que es la que se incluye de facto dentro de la formación del sociólogo.

---

<sup>22</sup> Las *suites de cómputo*, *software suites* hace referencia al programa, conjunto de programas o conjunto de aplicaciones que permiten crear, modificar, organizar archivos de texto, hojas de cálculo o procesar imágenes. Algunos de los tipos de suites existentes son las suites ofimáticas, ejemplo de ello tenemos *Microsoft Office*, *Open Office*, *Google G Suite* y *iWork* en el caso de Apple; las suites de procesamiento de datos como *Hadoop Apache*, *R*, *Stata*, *SPSS*, *Gaphi*, *Pajek*; de programación con *java*, *Python*, *SQL*, *VisualBasic*, *C++*, *PHP* y *Swift*; y

Cabe mencionar que nos referimos como ecosistema digital al conjunto de herramientas utilizadas para el monitoreo y recolección de datos, que al igual que su homólogo en biología funciona para especificar al conjunto de subsistemas que interactúan entre sí, dentro de un espacio determinado. Estos entornos digitales son generalmente utilizados por los medios de comunicación masiva y en diversas ocasiones por periodistas para administrar el flujo de datos sobre los acontecimientos que ocurren alrededor del mundo, por lo que la creación de estos entornos puede jugar un papel importante en favor de la sociología dentro de la investigación, debido a que su creación influye en la forma en que se recaban los datos, además del tratamiento de la información. Poniendo un ejemplo sobre esto, es actualmente con el tema del monitoreo de propagación de enfermedades, siendo más específico el *coronavirus* (COVID-19), ya que, con la información de las organizaciones de salud, la cobertura de los medios de comunicación y los sistemas de geolocalización se ha podido crear un ambiente adecuado para monitorear el avance del virus, así como el número de casos existentes (contagios y decesos).

Si bien, la creación de un ecosistema digital posibilita al investigador a tener una mayor perspectiva sobre el fenómeno que ésta siendo analizado, crear estos entornos propicios para recopilar la información resulta una práctica común para el sociólogo, ya que al formular nuevas investigaciones y plantear el objeto de estudio comenzamos a construir en nuestra mente una esfera en donde recolectaremos dicha información para así poder analizarla en un futuro con más detenimiento. Sin embargo, a pesar de establecer estas esferas de datos de manera instintiva, crearlas de manera consiente con la incorporación de variables relacionales en entornos digitales, resulta una tarea más complicada. Como se ha mencionado a lo largo del texto, una de las cualidades que debe considerar el científico social es mantener un control sobre los datos, sin llevarlos al extremo, ya sea estratificar demasiado la información o mejor dicho depurar tanto los datos que pierdan su valor (manipulación de los resultados de la muestra), o por el contrario, que sea tanta información sin organizar que no se pueda analizar debido a que los datos no se

---

por último las suites de diseño como Adobe Creative Suite, CorelDRAW Graphic Suite, Autodesk. Cabe resaltar que los propios sistemas operativos Windows, Android, IOS/MacOs cumplen la función de ser software suites.

encuentran en condiciones para ser tratados, que de igual forma crea un sesgo en los resultados. Por ello, los ecosistemas digitales más que para entender el funcionamiento que puede tener la información en Big Data como parte de la sociología, permite a los científicos sociales poder manipular toda la metadata recolectada que puede ser útil para el abordaje y el estudio del fenómeno social.

Veámoslo con el siguiente ejemplo, supongamos que un sociólogo pretende hacer un estudio comparativo sobre la eficiencia de los sistemas de transporte público en la ciudad. Para ello, debe consultar datos relacionados al tema como: los datos demográficos de la ciudad, los tipos de transporte público que existen en la localidad, el número de usuarios promedio y las vialidades existentes, dichos datos son conseguidos de manera tradicional, a través de las instituciones, por lo que se encuentran ligados al Open Data y conforman la primera parte en la construcción de la esfera de datos; sumando a ello, también deben ser considerados datos que nos permitan generar comparaciones como los tiempos de traslado, aspectos de seguridad, quejas de los usuarios sobre el transporte, las unidades que circulan, el número de usuarios en las aplicaciones, además de las aplicaciones de transporte que hay, y que formarán la segunda parte de la esfera de datos, la cual se encuentra relacionada con los otros dos niveles de Big Data (Social Data y Geo Data). Todo esto, para tener una mejor perspectiva sobre el problema o fenómeno social que estamos analizando y asimismo crear las condiciones idóneas para mantener un estado de vigilia constante o, en otras palabras, monitorear el flujo de información sobre nuestro objeto de estudio. Aunque cabe señalar que la construcción de estos ecosistemas digitales se encuentra sujeto a la interpretación o las necesidades de cada investigador, por lo que pueden existir elementos similares al momento de construir los entornos para recolectar la información, pero no serán los mismos en cada esfera de datos construida.

### **3.2 Recursos de la sociología en el Big Data**

Aunque el análisis de los grandes datos ha resultado en beneficio de múltiples disciplinas debido a la versatilidad con la que los repositorios de información pueden generar aportes mucho más precisos a las diversas problemáticas sociales. En el caso de la sociología ocurre algo peculiar, a pesar de que el análisis de datos pueda arrojar tanta información relevante como sea posible para nuestra investigación, no se debe perder de vista que, sin una fundamentación teórica de nuestras variables, los resultados obtenidos de esos análisis terminan convirtiéndose en cálculos relacionados a algún tema. Por lo cual, este apartado tiene la finalidad de mostrar los aportes que tiene la sociología en el abordaje de los fenómenos sociales con Big Data.

Como se ha podido observar en los capítulos anteriores, el Big Data cuenta con diversas herramientas que pueden facilitar, en la medida de lo posible, el trabajo del sociólogo al momento de analizar datos, refiriéndonos a los distintos niveles de Big Data, así como las características que lo conforman. Sin embargo, por muy bueno que pueda resultar cualquier análisis hecho con Big Data, la complejidad de ellos recae en la labor del sociólogo para descifrar esa información, por lo que se deben contrastar los resultados obtenidos con los referentes teóricos tomados para su estudio, ya que de esta forma se dejan ver los resultados de manera concreta, es decir, se revelan los cambios en la estructura social, la aparición de un factor determinante que incurra en el comportamiento social, o bien, que no existan modificaciones en las prácticas sociales.

En la actualidad, debido a la expansión de la industria digital, la sociología se ha visto rebasada en cierta forma por el rápido aumento de los datos, por lo que ha tenido que adaptarse a nuevas posibilidades de recopilar y analizar información, buscando el acceso a plataformas digitales que cuenten con vínculos institucionales, logrando fortalecer la interpretación de los datos, sumando además que los aportes enriquecidos por el análisis de extensos repositorios de información, combinados con las técnicas e instrumentos propios de la sociología, actúen de

forma complementaria para corroborar los hallazgos del análisis, dependiendo de cuál sea el objeto de estudio.

Por otro lado, el diseño de nuestro proyecto de investigación perfilado al análisis de Big Data es igual de importante que la recolección de los datos, ya que la estructura de nuestro proyecto dependiendo de cómo esté encaminada, permitirá contrastar e inclusive revelar patrones de comportamiento o nuevas estructuras sociales. Ahora bien, en cuanto a su diseño se debe pensar y estructurar esencialmente como cualquier otro proyecto de investigación, es decir que, debe contener objetivos principales y secundarios, preguntas de investigación, hipótesis (sí se requiere), marco histórico, marco teórico-metodológico, he aquí en donde el investigador deberá tomar en cuenta la inclusión del análisis de Big Data en su investigación y asimismo justificar la utilidad del análisis con los aspectos expuestos en el capítulo dos sobre *La importancia del analizar Big Data*, tomando en consideración las demás técnicas e instrumentos para recolectar la información<sup>23</sup> que formarán el entorno de control para el ingreso de datos.

Algunas de las técnicas e instrumentos más usadas por la sociología, y que de alguna forma puedan ser complementario del análisis de Big Data, se encuentra la encuesta, el cuestionario, la entrevista, los grupos focales en primer plano, debido a que estas tienen contacto directo con los informantes por lo que se pueden notar expresiones que virtualmente no se podrían notar, por otra parte, también se encuentran la observación participante que aunque puede parecer no ser de gran ayuda cuando de entornos digitales se trata, es todo lo contrario, es decir, la observación participante aplicada a un ambiente digital permite visualizar los comportamientos del individuo de manera más natural con la menor interferencia posible.

---

<sup>23</sup> Es necesario señalar que, este tipo de análisis si es adecuado para el objeto de estudio y si el investigador a cargo lo determina, debe ser incluido como parte de su instrumentación en la triangulación metodológica, siendo complementado por otros instrumentos que se adecuen a su investigación.

Tal vez el ejemplo más claro que tendríamos con la observación participante en ambientes digitales es precisamente con la interacción de los usuarios en plataformas digitales, como Reddit, Twitter o Facebook, es decir, plataformas/sitios web que se basan principalmente en compartir imágenes, noticias relevantes, foros/grupos, post sobre distintos temas (política, economía, educación, ciencia, teorías de conspiración, ficción, entre otros), de esta forma, a través de estos sitios es posible reconocer el contenido que puede ser de interés para los usuarios y poder parametrizarlos por medio de reacciones en Facebook, Retweets en Twitter, o flechas arriba en Reddit.

Esto mismo también puede ocurrir en el caso de las encuestas y cuestionarios aplicados a través de medios electrónicos, si bien, pueden ser una gran opción ya que extiende el alcance que puede tener la aplicación del instrumento, brindando mucha más información, también puede ser contraproducente para el investigador, debido a que si no se tienen puntos de control durante su aplicación o el instrumento no se encuentra diseñado correctamente puede sesgar los bancos de datos recolectados, haciendo que las interpretaciones del investigador sobre el comportamiento del individuo difieran de su hipótesis o pregunta inicial. No obstante, a pesar de que los instrumentos pueden ser aplicados digitalmente, algunos otros como las entrevistas o los grupos focales no pueden ser aplicados de esa forma, ya que estos dependen más de la percepción del investigador sobre los datos del informante o como mediador del grupo sobre los participantes. Sin embargo, eso no impide que posteriormente de la realización de un análisis de Big Data el investigador considere aplicar alguno de ellos para contrastar los resultados del análisis con los datos provistos de sus informantes.<sup>24</sup>

---

<sup>24</sup> Cabe resaltar que, en la medida de lo posible, se pueden hacer combinaciones de cualquier técnica o instrumento con el análisis de Big Data, ya que no por ser un análisis que involucre grandes repositorios de datos deberá ser tratado como un objeto ajeno al proyecto de investigación. Sumado a ello también, las técnicas e instrumentos antes mencionados son los utilizados recurrentemente en la disciplina, por lo que no reflejan el uso exclusivo de estas con el análisis de Big Data.

Ciertamente, aunque la sociología en sí misma permite describir las distintas estructuras dogmáticas arraigadas de la sociedad, y a su vez permite reconocer en la medida de lo posible el porqué de las acciones entre los individuos en contextos determinados. La incorporación del Big Data a las ciencias sociales puede resultar para el investigador de gran utilidad, debido a que no tan solo el análisis de los grandes datos se enfocaría en estudiar los fenómenos sociales o su estructura, sino también en estudiar a los sujetos que forman parte del fenómeno, ya que toma en cuenta los distintos factores que lo hacen relacionarse con los demás, refiriéndose a los grados de afinidad existente entre los distintos actores.

Por lo que, la presencia del sociólogo en el ambiente digital más allá de ser un usuario presente en un medio de comunicación alternativo funge como un observador que mantiene un estado de vigilia constante sobre los datos que circulan a su alrededor, y que a su vez discierne sobre la importancia que pueden tener para sus investigaciones, similar a lo que menciona Foucault con el panóptico y ejemplo de la prisión en donde el guardia apostado en la torre se encuentra vigilando continuamente el movimiento de los reclusos sin intervenir en su comportamiento. Resultando en interpretaciones que se enfocan más allá de los tópicos de investigación comunes como el comportamiento electoral o las clásicas estructuras paradigmáticas de la sociedad, sólo por mencionar algunos ejemplos.

### **3.3 La analítica de Big Data desde una perspectiva sociológica**

Tal vez hasta este punto, el lector ya se habrá percatado de la pericia que supondría tener el sociólogo para realizar un análisis de Big Data, considerando que tiene que adquirir de conocimiento básico en programación. No obstante, ahora que se han establecido los distintos niveles que posee Big Data (Open, Social y Geo Data), se han mencionado las fuentes de datos, además del potencial que tiene al incluirlo en las investigaciones ya no tan sólo de las ciencias sociales, sino también de la sociología, queda por mencionar tal vez un último aspecto que terminaría por estructurar la incorporación del Big Data como una herramienta útil para la

sociología; y esa es la parte instrumental de los grandes datos, refiriéndonos específicamente a la analítica presente en Big Data. Por lo que este apartado tiene la finalidad de conocer su definición y asimismo establecer su vínculo con la sociología.

Pero antes de continuar es necesario establecer ¿Qué es la analítica? Si bien el término hace referencia al análisis de cualquier conjunto de datos, y esta definición es válida, el término puede ser definido mejor como una serie de “procesos y actividades diseñados para obtener ayuda y evaluar datos” (ISACA en Joyanes 2013:240) que puedan ser de utilidad para explicar las condiciones de una problemática o fenómeno dentro de un tiempo o contexto determinado, ejemplo de ello, es el análisis estadístico y la demografía. Sin embargo, cuando se trata de la analítica en grandes volúmenes de información, puede resultar un poco más complejo que sólo su tratamiento, por lo que el análisis de datos dentro de Big Data, pese a que consiste en una serie de operaciones a las que se somete la información con el fin de interpretar conjuntos de datos estructurados, semi estructurados y no estructurados, también debe considerar los distintos tipos de analítica existente.

Por otra parte, debido a que la clasificación en la analítica suele variar entre autores, algunos siendo más técnicos en su clasificación y otros más utilizando categorías subjetivas que van de lo descriptivo hasta la construcción de escenarios predictivos, dependiendo del campo al que va dirigido (académico, empresarial, publicitario o político, entre otros), se pueden vislumbrar tres clases de analítica de Big Data, además de la analítica tradicional, y cuatro niveles de análisis,<sup>25</sup> que permiten comprender un poco mejor la metodología estructurada en los grandes datos para su incorporación en la sociología.

La primer clase de analítica que se puede encontrar es *la analítica de Big Data*, que, según la explicación señalada por Joyanes (2013), el término se encuentra compuesto por dos partes; analítica y Big Data, por lo que se podría definir como

---

<sup>25</sup> Llámese IBM, Oracle, Azure, Gartner o cualquier otro distribuidor de servicios digitales de cloud computing, hosting o similares, además de los académicos, tomando como referencia la clasificación de Joyanes para la distribución de los tipos de analítica.

“... el uso de técnicas analíticas aplicadas a conjuntos de grandes volúmenes de datos” (Joyanes, 2013: 243), con la intención de examinar velozmente una amplia variedad de información para visualizar patrones ocultos y correlaciones inusuales que permitan identificar áreas de interés. Es por ello que su utilización más frecuente es el área comercial y/o empresarial para valorar los factores de riesgo como el fraude, así como eficientar los procesos internos a través de modelos algorítmicos que contribuyen en la toma de decisiones dentro de una empresa y/o negocio; la segunda clase de analítica presente, es *la Analítica Web & Social*, que se basa principalmente en el análisis del tráfico de datos dentro de las redes sociales, buscadores web y plataformas digitales como aplicaciones, por lo que se utiliza frecuentemente en publicidad y estudios de mercado para perfilar poblaciones objetivo que puedan adquirir el producto; y por último, la tercer clase de analítica es *la Analítica móvil o Analítica binaria*,<sup>26</sup> que se basa principalmente en el análisis de datos proveniente de dispositivos GPS, Chips NFC y RFID, sensores y dispositivos IoT, por lo que es utilizada para el monitoreo de sistemas de mapeo y geolocalización, además de sistemas automatizados o semi automatizados que no requieran tanto de la intervención de los usuarios, es decir, casi cualquier conexión entre dispositivos digitales, refiriéndose a la comunicación de máquina a máquina.

Por otra parte, como ya se había mencionado, además de la analítica existen cuatro niveles de análisis, que dependen de la estructura del proyecto y los objetivos que tenga el investigador sobre los datos recolectados, por lo que, la interpretación del análisis puede tornarse descriptiva, diagnóstico, predictiva o prescriptiva. Pero ¿en qué consisten estos niveles de análisis que forman parte de Big Data? Bien, la respuesta radica en el entendimiento de dos cosas; la primera es que la analítica se basa en la examinación de los conjuntos de datos, por medio de operaciones que puedan llevar a un resultado óptimo, como ocurriría en el análisis estadístico. Cabe mencionar que entre mayor sea la calidad de los datos, puede arrojar una mejor interpretación de la realidad social hablando en términos sociológicos, lo cual enriquecería la investigación; lo segundo que debemos entender, es la complejidad

---

<sup>26</sup> Refiriéndose a la codificación binaria como base en el lenguaje computacional y por el cual hoy en día la mayoría de los dispositivos conectados entre sí usan un código numérico de comunicación

de los objetivos que tenga el proyecto, ya que de ellos dependerá en un primer momento del nivel de análisis que se requerirá para la interpretación de los datos que estamos examinando. He aquí donde también juega un papel importante por parte del investigador/sociólogo sobre el conocimiento contextual o temporal en que es realizado el análisis.<sup>27</sup>

Dicho de otro modo, las condiciones en las que se formula el análisis es lo que determina su complejidad, ya que la analítica al ser "...el mejor medio para descubrir nuevos segmentos de clientes, identificar los mejores proveedores, asociar productor por afinidad, comprender la estacionalidad de las ventas, etc." (Joyanes, 2013: 243), permite, además, comprender de una mejor forma el comportamiento social y con ello también la respuesta de interacción – acción – reacción del sujeto en un entorno natural, tomando como parte de ese entorno ya no tan sólo el físico, sino también el digital.

### **3.3.1 Descriptivo**

Ahora bien, continuando con la explicación, el análisis descriptivo se puede considerar como el primer nivel de análisis observable en Big Data, por lo cual también se puede considerar como el nivel con menor dificultad en cuanto al análisis y la interpretación de los datos, esto, debido a que al ser una etapa preliminar del procesamiento de datos, la mayoría de las operaciones realizadas se basan en la estadística descriptiva<sup>28</sup> y aritmética básica (sumatorias, promedios y porcentajes) y suele presentar la información mediante gráficos, cuadros, tablas, hojas de

---

<sup>27</sup> Nota: Si bien, los términos de analítica y análisis pueden ser equivalente e incluso funcionar como sinónimos en el texto; dentro del apartado se tomará como análisis a la técnica de procesamiento de datos, mientras que analítica como la metodología que interactúa con la técnica para el tratamiento de los datos.

<sup>28</sup> Recordando un poco, la estadística descriptiva, permite ordenar y clasificar conjuntos de datos de tal forma que sea posible trabajar con ellos. Por lo que suele recurrir a la representación gráfica de los datos mediante tablas o cuadros que permitan visualizar dicha información, además de establecer parámetros que ayuden a simplificar la complejidad de los conjuntos de datos que se hayan podido obtener de la aplicación de un instrumento, o bien, de otras bases de datos más extensas. Aunado a lo anterior, la estadística descriptiva en cierta forma se encuentra limitada a hacer deducciones directas con relación a los datos y los parámetros establecidos.

síntesis o resúmenes (véase figura12.), es decir, cualquier formato que pueda facilitar la lectura de los resultados.<sup>29</sup>

**Figura 12. Tipos de gráficos simples para el análisis descriptivo**



Fuente: Elaboración propia a partir de las imágenes de gráficos de Microsoft word

Como su nombre lo indica, este tipo de análisis es usado para describir escenarios inmediatos, o bien, de manera preliminar para ubicar el contexto del problema o fenómeno que se está investigado. Asimismo, es posible notar que el análisis descriptivo dentro de la sociología, son los primeros intentos del investigador por lograr interpretar la realidad social, ya que, si nos pusiéramos a reflexionar sobre las preguntas iniciales que haríamos para establecer las bases de nuestro análisis, probablemente lo primero que se nos vendría a la cabeza sería el ¿qué? ¿cuándo? y ¿dónde?, hasta este punto es algo parecido a la construcción del estado del arte

<sup>29</sup> Como alternativa en la forma de presentar los datos, pueden ser utilizadas infografías, ya que, debido al diseño compacto para presentar la información, además de que en su mayoría los datos pueden ser acompañado de imágenes que enriquezcan en texto.

cuando planteamos un problema, es decir, en ambos casos tratamos de establecer el contexto apropiado del fenómeno a analizar, pero con la diferencia de que en el análisis descriptivo los datos ya se encuentran agrupados, simplificados y óptimos para ser presentados, junto con nuestras otras variables, contemplando responder las preguntas iniciales; lo que en el estado del arte aun debemos clarificar conociendo cuales han sido los abordajes más recientes. Sin embargo, no se debe perder de vista que también el análisis descriptivo puede presentarse como la etapa de recolección de datos durante la investigación social o lo que es lo mismo la etapa de experimentación siguiendo el método científico.

En otras palabras, el sociólogo durante su formación recurre a la exploración de bases de datos prediseñadas<sup>30</sup> con la intención de aprender a desenmarañar la información que se encuentra en esos repositorios, y así establecer de manera óptima un contexto histórico, económico, político o social. De tal manera que se comienzan a formar interpretaciones sobre el comportamiento social de los individuos a través de datos que pueden ser mayormente cuantitativos en los primeros acercamientos del investigador o cualitativos cuando adquiere mayor experiencia.

### **3.3.2 Diagnóstico**

Por otra parte, el análisis de diagnóstico, como su nombre lo indica, es un análisis más detallado de un problema o fenómeno, esto quiere decir que no tan solo se enfoca en agrupar y presentar la información como en el descriptivo, sino que pretende ir más allá, dando a conocer las causas, a través de una interpretación más profunda de los datos. Es decir, además del contexto inmediato hace referencia también a los antecedentes históricos, políticos y sociales que los originaron, aunado también al descubrimiento de variables que puedan permitir comprender de

---

<sup>30</sup> Cabe mencionar que algunas de las bases de datos prediseñadas, además de las que son creadas de manera ficticia, son las provenientes de instituciones gubernamentales como INEGI con los censos de población, CONEVAL con el registro del nivel socioeconómico organizado por zona territorial, INE con la distribución de votantes por distrito electoral a nivel local y federal, sólo por mencionar algunas.

mejor forma el fenómeno, por lo que además de las operaciones básicas (sumatorias, media, mediana, moda,) es necesario incorporar elementos propios de la estadística inferencial<sup>31</sup> como las desviaciones medias y estándar, las pruebas de hipótesis, los cruces y correlación de variables (solo por mencionar algunos elementos). Todo esto, con la finalidad de deducir características desconocidas de los conjuntos de datos ya conocidos.

Ahora bien, al igual que sucede con el análisis descriptivo en donde se recurre a formular preguntas en torno al contexto inmediato (¿cómo?, ¿cuándo? y ¿dónde?), en el caso del análisis de diagnóstico esas preguntas son insuficientes para satisfacer nuestras dudas, por lo que es necesario añadir otras que se adecuen con lo que requerimos. De esta forma, dentro de la lógica del investigador por conocer las causas/antecedentes del problema las preguntas que mejor se adecuan son ¿por qué? y ¿para qué? esto, con la intención de observar más a profundidad el origen del fenómeno, que en términos sociológicos se refleja en un mejor abordaje y una justificación más clara sobre lo que estamos investigando.

En cuanto a la aplicación de este tipo de análisis puede ser factible para ocasiones en donde ya se hizo un análisis exploratorio, pero aún no se tiene un contexto claro del origen del fenómeno y es necesario saber más de él, o bien, cuando se requiere un análisis más detallado de repositorios de información preconstruidos (como lo suelen ser las bases de datos gubernamentales). Cabe mencionar que el manejo de los datos de Big Data en sus dos primeras etapas no se aleja demasiado al que estamos acostumbrados en sociología, por lo que calcular algunos parámetros que nos puedan ayudar a vislumbrar un poco más el fenómeno se basa principalmente en catalogar y seleccionar la información que nos convenga a nosotros como

---

<sup>31</sup> La estadística inferencial, permite realizar interpretaciones más complejas a partir de los conjuntos de datos (como lo pueden ser las bases de datos prediseñadas) con la intención de descubrir variables desconocidas sobre el problema, asimismo, como examinar y obtener información que permita estimar las características de una población a partir de un subconjunto de datos extraído del conjunto principal. es decir, la selección de una muestra estadística en la que se realizaran cálculos, estimaciones que permitan interpretar de mejor forma el problema o fenómeno investigado. Además de ampliar nuestro campo de acción sobre el mismo. Cabe mencionar que la base de la estadística inferencial se encuentra en la estadística descriptiva, debido a que las operaciones que determinan los siguientes pasos dentro del análisis, son hechas con la estadística descriptiva.

investigadores. Claro está, sin descuidar el tratamiento y el control de los repositorios de datos, esto último principalmente para no generar sesgos innecesarios durante la investigación.

### **3.3.3 Predictivo**

El análisis predictivo como su nombre lo indica, permite realizar estimaciones sobre acontecimientos futuros a través de la construcción de modelos matemáticos y/o algoritmos que ayuden a conocer de mejor forma el desarrollo de un problema. Este tipo de análisis suele ser usado principalmente con fines comerciales para evaluar riesgos de inversión, cambios en la fluctuación del mercado, estimar el crecimiento y el alcance de una empresa; además de tener aplicaciones en la industria tecnológica en sus ramas de inteligencia artificial y automatización de procesos de los que se puede destacar los dispositivos médicos para prevenir enfermedades o suministrar medicamentos, hasta la industria aeroespacial. Es decir, el análisis predictivo es una técnica que combina la estadística aplicada<sup>32</sup> y la probabilidad estadística<sup>33</sup> con la interpretación de datos cuantitativos a través de modelos matemáticos.

Ahora bien, aunque los niveles anteriores nos recuerdan un poco al proceso de investigación en sociología, por las similitudes que se tienen con el análisis descriptivo y diagnóstico con la presentación de información, como en el establecimiento de los antecedentes y origen del problema/fenómeno social. En el caso del análisis predictivo estas similitudes ya no son tan evidentes, considerando que el tratamiento de los datos se basa principalmente en la aplicación de procesos estadísticos relativamente más complejos para lograr obtener modelos matemáticos que capturen las tendencias importantes con el fin de obtener resultados óptimos.

---

<sup>32</sup> La estadística aplicada está conformada por las dos clases de estadísticas anteriores. Su objetivo consiste en deducir resultados sobre un universo, a partir de una muestra determinada. Este tipo de estadística puede ser aplicada en cualquier área que no pertenezca a ella, tal como historia, psicología, etc.

<sup>33</sup> La probabilidad estadística es una forma de medición de la certidumbre que asociada a la observación u ocurrencia de un fenómeno o al hecho de que una característica de un objeto de estudio adopte cierto valor. Se puede simplificar dividiendo el número de ocurrencias de un hecho entre el número total de casos posibles.

Visto desde una perspectiva más enfocada a las ciencias sociales y a la sociología se podría notar que el análisis predictivo guarda cierto parecido con otra técnica más cercana a la disciplina como lo es el análisis de coyuntura, debido a que el objetivo principal de ambas radica en la construcción de escenarios probables de un problema o fenómeno, por lo que el análisis predictivo al igual que el análisis de coyuntura suele hacer uso de conjuntos de datos de tipo histórico, económico, social, político o cultural para establecer patrones de comportamiento que le permita interpretar la evolución del problema o el desarrollo del fenómeno a través de distintos entornos. Sin embargo, como recordaran al principio del texto, Big Data no se describe solamente como un conjunto de datos masivos que pueden ser estructurados, semiestructurados y no estructurados, sino que también cuenta con técnicas propias para la extracción y análisis de los datos, mismas que pueden funcionar en algunas ocasiones como técnicas complementarias a los tipos de análisis, siendo una de esas técnicas la minería de datos, que en la mayoría de las ocasiones -si no es que en todas- cuando se menciona la realización de un análisis predictivo sobre cualquier tema estas dos se encuentran a disposición una de la otra. Es decir, por una parte, la minería de datos nos permite explorar más fácilmente los grandes repositorios de información para encontrar factores comunes que puedan percibirse como un patrón algo que se conoce normalmente como extracción de conocimiento en bases de datos; y, por otro lado, la utilización de la estadística aplicada, la probabilidad estadística y los datos cualitativos para interpretar esos patrones y poder predecir así el comportamiento social.

En otras palabras, "...el muestreo estadístico propio de tiempos de escasez de información estaría pasando a la historia cuando ésta es abundante" (Meneses, 2018: 424), y es aquí donde el análisis predictivo que se realiza en Big Data se vuelve relevante también para la sociología, debido que en ciertas ocasiones por más ávido que sea el investigador, este llega a tener problemas para encontrar esas conexiones lógicas dentro de las bases de datos que le permitirán realizar una mejor interpretación del comportamiento social. Es decir, al igual que sucede con los *puzzles* en donde la solución se va revelando en la medida que se va interactuando con sus elementos, con la investigación suele pasar algo similar ya que al tratar de

establecer cruces o conexiones entre las distintas variables notaríamos que algunos elementos no podrían encajar, pero que aplicando minería de datos y el análisis predictivo el número de variables relacionadas con nuestro objeto de estudio aumentaría, revelando también algunas nuevas que contribuyan a la interpretación del fenómeno social. Por lo que, de alguna manera sumar esta posibilidad como parte del repertorio de técnicas de la sociología capacitaría al investigador para reforzar sus teorías y aumentar el valor de sus resultados.

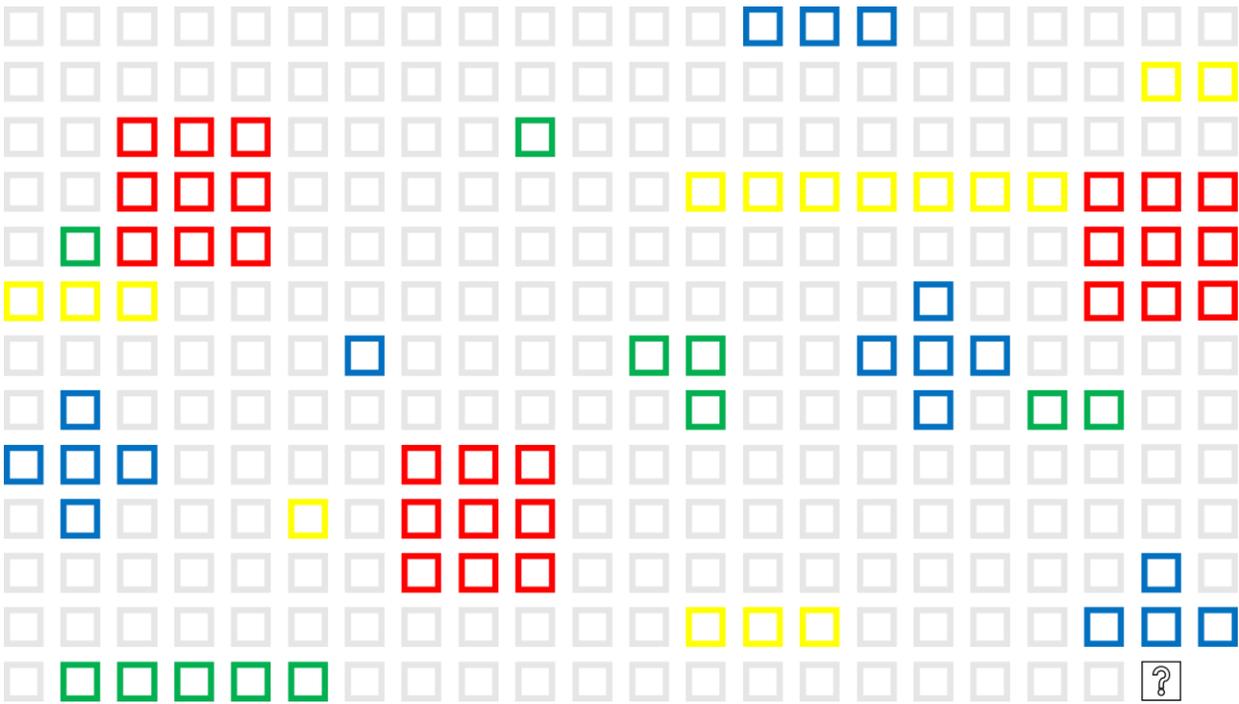
Siguiendo un poco su composición dentro de la investigación sociológica encontraríamos al igual que en los anteriores, que este tipo de análisis responde la pregunta ¿qué utilidad tiene el dato para nuestra investigación? Obviando que es necesario para la fundamentación del trabajo o que sin ellos no se puede continuar con la investigación, es decir, como investigadores debemos ser conscientes que al tratar con amplios repositorios de información estos pueden llegar a modificarse cambiando también nuestras necesidades, algo similar a lo que sucede con los datos cualitativos por lo que la relación del sujeto-objeto puede ser modificada en función de los testimonios, el comportamiento y la interacción con el informante.

Una forma sencilla de entender cómo funciona el análisis predictivo en general y particularmente en sociología es de la siguiente forma, imaginemos que un investigador quiere conocer los hábitos de comportamiento de personas de 25 a 30 años en redes sociales, y para ello, utilizará la minería de datos para recolectar la información de publicaciones de Facebook y Twitter, ahora bien, cada uno de estos se irán almacenando en forma de ficheros que se analizarán en busca de información que pueda coincidir. Cabe mencionar que debemos tener en consideración que este tipo de análisis se basa en la explotación de ficheros de datos con el fin de encontrar información útil. Por lo que, estas coincidencias pueden ayudar a vislumbrar el contexto, el comportamiento, las preferencias e inclusive deducir si el siguiente fichero puede contener parte de ese dato.

Como se puede observar (véase figura 13.), el diagrama anterior es una representación de cómo se podría ver un análisis predictivo usando la minería de datos si se mostrara de manera gráfica en una pantalla, en el cual cada uno de los

cuadros simbolizaría un conjunto de datos dentro de un repositorio de información, y a su vez que puede ser parte o no de un mismo tipo de dato, por lo que, al ir minando, explorando y analizando la información contenida en ellos, es posible notar la formación de patrones como los cuadros rojos y las cruces azules. Sin embargo, para que estos patrones sean visibles y no sean un producto resultado por mera coincidencia son necesario poder predecir la secuencia de los datos siguientes como se muestra en el diagrama con la cruz azul, es decir, que conociendo un poco el comportamiento de nuestras variables gracias a que se cuenta con un mayor número de datos para analizar se puede predecir el comportamiento de los datos con mayor facilidad.

**Figura 13. Representación gráfica sobre cómo se vería un análisis predictivo utilizando minería de datos**



Fuente: Elaboración propia

### **3.3.4 Prescriptivo**

A grandes rasgos la analítica prescriptiva es un sistema que recopila y almacena información con la intención de clasificar, seleccionar y comparar datos entre distintos ficheros para determinar los escenarios más favorables que contribuyan a la toma de decisiones sobre un problema. En otras palabras, analiza el impacto que pueden tener las leyes en una población/comunidad, la estabilidad de un sistema económico, o bien, el crecimiento de una empresa/negocio mediante la integración del análisis descriptivo y el análisis predictivo.

En la mayoría de las ocasiones se pueden ver este tipo de sistemas aplicados en empresas o negocios para eficientar los procesos administrativos, además de establecer estrategias que puedan ayudar a mejorar a la empresa mediante el análisis de ventas, opiniones positivas o negativas, tamaño del inventario, métricas en el rendimiento de los empleados y las sucursales/centros operativos. Uno de los ejemplos más claros que podríamos tener sobre el funcionamiento del análisis prescriptivo es con Amazon, ya que al evaluar los parámetros indicados anteriormente crea recomendaciones favorables para los clientes y de esta forma generar más ventas; otro ejemplo de esto se logra notar con las cadenas hoteleras y los restaurantes de comida rápida.

Sin embargo, pese a que tiene su relación con los niveles anteriores que resultan ser más sencillos en su incorporación de la sociología, este último tipo de análisis no resulta ser tan simple debido a que está compuesto "...por una colección de técnicas relacionadas y tipos de herramientas que normalmente incluyen analítica predictiva, minería de datos, análisis estadísticos y programación compleja de SQL" (Joyanes, 2013: 243), de manera que al ser un tipo de análisis que incorpora todos los demás niveles es necesario la optimización de recursos dentro de un clúster concentrado en un servidor propio o de pago, por lo que pocas veces se podrá utilizar en una investigación, pero que no es imposible de ver, ya que los usos comunes que se tiene, además de los comerciales, son relacionados a los procesos electorales, campo disciplinario en el que podemos observar también a la sociología. Ahora bien, estas complicaciones se deben principalmente por dos

cosas; la primera, es que el investigador debe contar con conocimientos avanzados de programación<sup>34</sup>, además de una formación estadística plena que permita desarrollar modelos estadísticos que le sean útiles para desarrollar su investigación; el segundo, como se ha venido mencionando, es el poder de cómputo y que se refiere a la implementación de servidores.

No obstante, pese a que el análisis prescriptivo es de los más complejos, al menos desde el punto de vista sociológico, debido al conocimiento especializado que requiere y por el poder de cómputo que se necesita, puede ser uno de los más adecuados para estimar el comportamiento y desarrollo de nuestro objeto de estudio, fenómeno social o problema en general, ya que permite comparar distintos escenarios para determinar el más favorable. Por último, cabe señalar que lo anterior no exime la utilización del análisis en casos específicos y si la ocasión lo requiere.

---

<sup>34</sup> Se debe contar con una formación en mantenimiento de servidores o tener una formación relacionada con sistemas de cómputo, bases de datos, además de estadística computacional. Debido a que el análisis prescriptivo requiere de la implementación de redes neuronales para realizar cálculos de variables en tiempo real y a gran escala.

## **Conclusiones. Ventajas y desventajas del Big Data en la sociología**

Resumiendo hasta ahora, podemos decir que el Big Data es un cúmulo de datos en bruto que son generados a través del envío y recepción de información de dispositivos móviles, computadoras, dispositivos IoT (internet of things) o cualquier dispositivo en general que pueda conectarse a internet, los cuales pueden ser almacenados, clusterizados y procesados, mediante los diversos niveles de análisis y técnicas como la minería de datos o el ARS, que en combinación con la sociología permitiría vislumbrar de mejor forma el comportamiento social. Es decir, lograr establecer nuevos enfoques a los paradigmas clásicos de la sociología e inclusive tender nuevos campos de estudio que no han sido desarrollados propiamente por la disciplina, pero que se encuentran presentes día a día con el uso de las redes sociales o cuando navegamos por el ciberespacio en busca de información, debido a que dejamos un rastro digital que se encuentra ligado a una identidad virtual desarrollada por nosotros y que dotamos de una personalidad. En otras palabras, el surgimiento de una nueva rama en la sociología digital o la evolución de esta vertiente de la sociología.

Sin embargo, no todo es miel sobre hojuelas cuando se trata de establecer un vínculo entre el Big Data y la sociología, por lo que se debe tener en cuenta ciertas consideraciones, y es que, dada la naturaleza del Big Data para gestionar grandes cantidades de datos, puede traer consigo algunos pros y contras que el propio investigador debe considerar al momento de considerarlo parte de su investigación. Es por ello, que este apartado tiene la finalidad de explicar las ventajas y las desventajas que se tienen al incorporar un análisis de grandes datos como parte de una investigación de carácter sociológico. Cabe mencionar que las ventajas y desventajas que se presentan a continuación son hechas desde un punto de vista reflexivo a partir de la información recabada para este trabajo, por lo que el investigador puede considerar otras que aquí no se plasmen.

Visto desde la perspectiva sociológica, se podría creer que son más las desventajas que las ventajas que traería consigo aplicar un análisis de Big Data. Pero, lo cierto es que los diversos campos de aplicación han hecho que la idea se pueda concebir como parte del repertorio de instrumentos e incluso técnicas de la sociología. Es por ello, que nos centraremos en mencionar las ventajas y desventajas más puntuales que traería consigo el Big Data a la sociología, y cómo afectarían nuestro trabajo como investigadores.

### **Mayor flujo de información**

Una de las principales ventajas que se puede encontrar relacionada con Big Data es el flujo de información, ya que por la naturaleza del Big Data, tener un mayor caudal de datos a nuestra disposición nos permite ampliar nuestro universo muestral tanto como lo necesitemos, por lo que ya no estaríamos limitados físicamente al número de encuestas que podemos levantar en un día o al número de ficheros que se pueden revisar a la semana, lo que permitiría analizar con mayor eficiencia el comportamiento de nuestras variables. En otras palabras, optimizaría considerablemente la sistematización de información encontrando relaciones entre variables que no eran visibles a simple vista pero que aumentaría la calidad de los datos, y por ende también se tendría resultados más precisos a corto, mediano o largo plazo, dependiendo del enfoque y el abordaje del investigador. Aunado también a que un mayor flujo de información nos permitiría explorar en otras áreas de conocimiento haciendo que la relación entre lo cuantitativo y lo cualitativo sea mucho más estrecha.

Aunque un mayor flujo de información dentro de Big Data puede traer consigo ventajas para la sociología como la mejora en los resultados a corto, mediano y largo plazo, que permitan discernir de mejor forma las características de un fenómeno, también se puede convertir en una desventaja, siendo que al tener una mayor cantidad de datos siendo recolectada, almacenada y categorizada, es fácil que se pueda perder en control de la muestra, lo que podría suponer un problema

al momento de analizar la información, ya que se convertiría en una incesante lucha del investigador al no poder establecer los parámetros adecuados para clasificar y seleccionar los datos más relevantes, trayendo consigo la creación de sesgos de información involuntarios o la falta de confiabilidad en los resultados haciendo que se vuelva inviable la investigación.

Algunas consideraciones que podríamos tener al momento de trabajar con grandes repositorios de información es establecer en la medida de lo posible variables de control, esto permitirá mantener el control al seleccionar los datos, pero a su vez nos permitirá ir clasificando la información e iniciar con la categorización, lo que mejoraría el flujo de los datos además de que posteriormente nos permitirá analizar de mejor forma los datos que recolectamos. Por ejemplo, algunas de estas variables de control pueden ser el público al que está dirigido el estudio, las características de nuestra población, nivel socioeconómico, hasta objetivos principal y secundarios sirven de apoyo para mantener vigilada nuestro levantamiento de información. Otra consideración que debemos tener para no perder de vista el objetivo de nuestra investigación cuando contamos con un mayor flujo de información es tratar de estratificar nuestra información, es decir, agrupar los datos en clústeres lo suficientemente grandes para que puedan ser analizados de forma independiente pero que formen parte de un mismo grupo.<sup>35</sup> De esta manera, facilitaría el análisis general de todo nuestro repositorio de información.

## **Versatilidad**

Si bien, Big Data nos permite consultar y analizar variadas fuentes de datos en un periodo corto de tiempo gracias a que posee un mayor caudal de datos. La segunda ventaja que se destaca en él es la versatilidad, que en principio deberemos entender

---

<sup>35</sup> Un ejemplo tal vez un tanto burdo de esto, sería el problema del camión de manzanas, en dónde el dueño de una frutería recibe un cargamento de manzanas de distintos tipos, rojas, verdes, amarillas y pintas todas revueltas, por lo que debe organizarlas en cajas más pequeñas. Es decir, las manzanas representarían las distintas clases de datos que han sido cosechados, el camión representaría nuestro clúster central o almacén donde se guardaron los datos y por último las cajas representarían los clústeres más pequeños donde el analista (dueño de la frutería) clasificaría los datos (las manzanas).

como la combinación de la adaptabilidad y la escalabilidad, refiriéndose al ajuste de parámetros hecho por el analista/investigador para explorar los datos que contienen los ficheros, sumando a ello también que podemos personalizar los parámetros del análisis, lo que nos permitirá ir incrementando el universo muestral cuanto se necesite para así obtener datos más precisos. Dicho de otro modo, los clústeres se pueden adecuar al tamaño del proyecto y al investigador, lo cual no tan sólo se reflejaría la utilización del Big Data en las disciplinas relacionadas con las ciencias duras, sino también ampliaría su compatibilidad con distintas disciplinas como la sociología, por lo que podríamos afirmar que parte de la versatilidad que tiene Big Data recae en que es modular, ya que tomaría los datos más relevantes o representativos del macro análisis y los incorporaría a la investigación para sustentar, falsear o responder las hipótesis o preguntas iniciales.

Pese a todas las ventajas que pueda traer consigo la versatilidad del Big Data en la sociología, también trae consigo varios obstáculos que debemos contemplar si es que queremos integrarlo como parte de nuestro repertorio de técnicas de análisis, dado que, por mucho que Big Data sea flexible, esa misma flexibilidad hace que nuestro análisis se vuelva inestable, Es decir, al ampliar, reducir o seccionar nuestras bases de datos, es posible provocar una ligera pérdida de datos que podría afectar en el análisis, por lo que debemos ser cuidadosos al manipular los datos, es decir, si bien, el Big Data tiende a ser volátil por naturaleza, esta inestabilidad puede afectar solo cuando el propio investigador/analista tiende a extremar el control de los datos o por el contrario hay ausencia de este al recolectar la información.

La mejor recomendación que se puede hacer aquí es estructurar plan de análisis que se adecue a la investigación, de esta manera se evitará en la medida de lo posible que se tenga que reducir la muestra o bien seccionarla, asimismo establecer parámetro o variables de control que permitan monitorear la muestra, reduciendo considerablemente la volatilidad en el flujo de información.

## **Proyección de la sociología en los análisis de Big Data**

Por otro lado, además de la flexibilidad que tiene Big Data para complementar el trabajo de investigación, se encuentra una ventaja que para algunos puede ser sutil, pero que a medida que los grandes datos se vuelven parte de la vida diaria, se logra percibir con mayor notoriedad la proyección que puede tener la sociología en Big Data, debido a que en el análisis de extensos repositorios de información donde predomina la parte estadística y matemática surge la necesidad de interpretar no tan sólo el comportamiento social, sino también la realidad social de los individuos que forman parte del análisis. Es allí en donde la sociología juega un papel como salvoconducto en la interpretación de los datos.

Por lo que mencionar la existencia o la creación de una relación entre ambas pueda ser favorable no suena tan descabellado, ya que como se ha mencionado, para la sociología el Big Data optimizaría los tiempos de recolección, clasificación y análisis de datos, además de que toda la maquinaria con la que cuenta Big Data para enfocarse en examinar, cotejar o explorar mayores volúmenes de datos que en ocasiones pueden afectar la extensión de la muestra, ya que en la medida de lo posible adoptaríamos técnicas de recolección de datos que son propias del Big Data, pero que estarían amalgamadas para utilizarse en la sociología; y por otro lado, la sociología le brindaría soporte teórico-metodológico, aportando perfiles más apegados al comportamiento social, pero también justificar de mejor forma la interacción entre individuos. Lo que para ambas partes haría que el valor no tan solo de los datos -que por sí solos ya tienen valor- sino también el valor intrínseco de nuestra investigación

Cabe mencionar que por mucho que automaticemos el proceso de análisis, en muchas de las ocasiones tendremos que hacer trabajo manual para analizar la información y corroborar los resultados, por lo que debemos ser conscientes que el análisis de Big Data no sustituye por completo el análisis del sociólogo, sólo lo complementa.

Tal vez el obstáculo más notable que puede tener este punto es justamente que para utilizar Big Data y poder incorporarlo como parte de la sociología, se deben tomar en cuenta que es necesario aprender programación, al menos en Python y SQL ya que son los utilizados más comúnmente en las distribuciones de software de cálculo o lenguaje matemático, asimismo como tener conocimientos avanzados en estadística aplicada, que en el caso de la sociología puede no ser un problema ya que es parte de la formación básica. Todo esto, con la finalidad de conocer el funcionamiento y la administración de las bases de datos para gestionarlos con mayor eficiencia, pero, además realizar los cálculos necesarios sin dificultad para estimar los parámetros de nuestro análisis, lo que puede causar el disgusto de algunos investigadores interesados en Big Data debido al tiempo que deben invertir.

### **Promueve la ampliación del conocimiento en áreas que no son propias de la sociología**

Considerar esta ventaja tal vez pueda sonar algo subjetiva para quien lo esté leyendo o para aquellas personas que buscan el pretexto adecuado para iniciar una especialidad referida a la ciencia de datos, sin embargo, en principio hay que reflexionar que la integración del Big Data como parte de las técnicas de análisis en la sociología puede ser un estímulo para la comunidad sociológica que constantemente intentan estudiar áreas de interés poco exploradas, pero que por la falta de recursos no han logrado aterrizar los proyectos.

En cuanto a la curva de aprendizaje que tiene Big Data para los nuevos usuarios es necesario mencionar que esta dependerá de las propias capacidades del usuario por adaptar o adoptar esos conocimientos como parte de su área de estudio. De esta forma, se podría decir que conocer y aprender de otras áreas de conocimiento como la informática, estadística aplicada a sistemas computacionales complejos o inclusive acceder a aprender un lenguaje de programación en forma alternativa a nuestra formación académica, nos permitiría asimilar de mejor forma la manera en que recolectamos la información, y, por ende, obtener un mejor análisis del

comportamiento social, y aunque ciertamente construir un proyecto de investigación con este tipo de análisis puede ser un área de nicho por las implicaciones que tiene, igualmente puede resultar motivante para aquellos que vienen de otras áreas con la intención de conseguir un enfoque distinto como sería con la sociología.

Desafortunadamente el tiempo es uno de los factores que puede converger en contra, respecto a esta ventaja y aunque si bien es cierto que gran parte de la comunidad sociológica opta por ampliar sus áreas de conocimiento mediante diplomados, especializaciones o estudios de posgrado, introducirse a áreas que no son propias de las ciencias sociales o que no tengan afinidad con la sociología puede ser complicado por el tiempo de inversión necesario para adquirir los conocimientos requeridos en el área elegida, como lo es en el caso del Big Data o su afín la ciencia de datos. Ahora bien, visto desde la perspectiva del sociólogo precisaría entre dos y cuatro años aproximadamente para adquirir el conocimiento necesario para realizar de manera óptima un análisis de Big Data.

Cabe mencionar que este periodo de aprendizaje puede variar por diversos motivos como, la curva de aprendizaje del interesado, sus habilidades y en cierta manera también su formación académica inicial, lo que podría llegar a causar desmotivación para algunos sociólogos, ya que la mayoría de los interesados en Big Data provienen de carreras afines como ingeniería en computación, informática, matemáticas aplicadas, sólo por mencionar algunos ejemplos. Con esto, no quiere decir que el Big Data signifique una pérdida de tiempo y recursos para el sociólogo, sino puede ser un punto de partida para realizar análisis complementarios dentro de la disciplina.

Realmente pocas son las recomendaciones que se pueden hacer en este punto, ya que la decisión de utilizar Big Data como parte de la sociología se queda sujeta a cada investigador y colega. Por lo que sólo puedo agregar que, si el interés en el área es legítimo, la mejor recomendación es realizar una especialización en ciencia de datos, de esta manera no se encontraría tan limitado al aplicar un análisis de Big Data.

## **Es posible trabajar desde la comodidad de un portátil o computadora de escritorio**

Por el contrario, a lo que nos podamos imaginar que para trabajar con Big Data es necesario contar con todo un laboratorio de cómputo al puro estilo cyberpunk con cuartos enteros llenos de torres que se la pasan computando 24/7 para analizar datos montado por nosotros mismos, lo cierto es que con la reciente llegada o comercialización de servicios de hosting y cloud computing, sustituyendo esa necesidad obligada del equipo de investigadores por invertir en sus propios servidores, por la posibilidad de contratar espacios de almacenamiento virtual que cumplen con la misma función. Con esto no quiere decir que el cloud computing sean un sustituto de los data warehouse, sabiendo que es el caudal del que se alimenta Big Data, sino quiere decir que las propias empresas dueñas de esos almacenes de datos (llámese Microsoft, Amazon, Google, IBM, Oracle, entre otras), permiten a terceros que puedan utilizar sus instalaciones para almacenar información y procesar datos que de otra forma no sería posible almacenar en un disco duro. Es decir, en términos más simples, facilita al investigador (es) para trabajar con una mayor cantidad de datos en equipos que pueden estar contenidos en un escritorio dentro de una oficina, ya que los accesos a la información son remotos por lo que contar con una conexión rápida de internet y un equipo de cómputo con una buena capacidad de procesamiento<sup>36</sup> debería bastar para procesar la información sin la necesidad de recurrir a soluciones más industriales.

Por pretencioso que pueda parecer esto, con el avance tecnológico actualmente es posible escribir, procesar datos y buscar información desde casi cualquier dispositivo que pueda conectarse a internet, por lo que la contratación de servicios de almacenamiento remoto puede ser factible para mejorar la administración de las bases de datos en flujos de información en Big Data. Sin embargo, la adquisición de servicios de cloud computing y hosting también puede significar una desventaja para los propios investigadores, ya que depende del costo-almacenamiento. Es

---

<sup>36</sup> Tal vez se tendría que hacer la aclaración aquí sobre el poder de cómputo del equipo, ya que no se trata de tener un equipo con componentes de última generación, sin embargo, entre más recientes sean los componentes es mejor el rendimiento del respecto al manejo virtual de los datos.

decir, por una parte, el investigador debe determinar la cantidad de espacio requerido para la información que contendrán sus bases de datos, lo que se puede traducir como el número de GB/TB/PB, además debe contemplar el tipo de dato (estructurado, semi estructurado, no estructurado) y la frecuencia de acceso a ellos; por otra parte, también debemos considerar el costo que implica tener esa cantidad de GB/TB/PB en la nube y es aquí donde se pueden encontrar con algunas complicaciones, ya que los costos están tasados en función del proveedor y del almacenamiento, por lo que pueden variar.

Para el sociólogo, esto puede significar una limitante en cuanto al manejo y tratamiento de la información, debido a que, independientemente del software que utilicemos para analizar, los datos nos encontraríamos en la necesidad de adquirir servicios de cloud computing y hosting. Lo que para algunos puede significar en un esfuerzo infructuoso ya que esta inversión se sumaría también a la inversión en el tiempo de formación complementaria para adentrarse al mundo del Big Data desde una perspectiva sociológica.

Aunque pudiesen parecer más los contras que los pros, al menos en este último punto sobre los servicios de cloud computing es solo una herramienta complementaria que nos ayudará a mejorar el rendimiento en cuanto al procesamiento de los datos, de manera que no sustituye nuestro trabajo de investigación, sino lo refuerza. Ahora bien, es posible que para aquellos interesados en Big Data se les dificulte comenzar administrando bases de datos de +1TB o inclusive conseguir los repositorios adecuados para aprender a manipularlos, por lo que es probable que comiencen con bases de datos de no más 25GB, y es aquí, en donde los costos de almacenamiento en la nube pueden reducirse considerablemente ya que algunos proveedores brindan cuentas con un espacio limitado de forma gratuita o bien a una fracción mucho menor del costo mensual por los gigabyte ocupados, siendo una opción bastante factible para aquellos sociólogos interesados en aprender sobre Big Data.

**Figura 14. lista de precios para servicios storage de IBM**

	Smart Tier		Standard	Vault	Cold Vault
	Variable		Previsible	Previsible	Previsible
	Coste automático optimizado para datos de cualquier actividad o acceso		Datos activos a los que se accede frecuentemente	Datos menos activos a los que se accede una vez al mes	Datos fríos a los que se accede pocas veces al año
Uso mensual de Smart Tier		Uso mensual de			
Capacidad de almacenamiento (USD/GB/mes)		Capacidad de almacenamiento (USD/GB/mes)			
Smart Tier - Hot ⓘ	0,0227 US\$	0-499.9	0,0238 US\$	0,0144 US\$	0,0072 US\$
Smart Tier - Cool ⓘ	0,0144 US\$	500+ TB	0,0228 US\$	0,0120 US\$	0,0072 US\$
Smart Tier - Cold ⓘ	0,0090 US\$	---	---	---	---

Fuente. Estimación de costos para servicios de hosting de IBM, Cloud Object Storage, disponible en <https://cloud.ibm.com/objectstorage/create#pricing>

### **Gestión ética en la recolección de datos para su utilización como parte de estudios de carácter académico y empresarial**

Subjetivamente este último punto puede ser considerado como ventaja y desventaja, debido a que, por una parte, con el uso de técnicas como la minería de datos proporcionaría mayor eficiencia en la extracción de información mejorando los procesos de recolección y clasificación de datos, lo que en términos generales se reflejaría en información de mayor calidad. Sin embargo, aunque el proceso de recolección de datos por sí solo puede ser complicado, ya que depende principalmente de los flujos de datos que tengamos disponibles, esta no es la desventaja de la que nos tenemos que preocupar esencialmente, sino de un aspecto que de no tomarse en cuenta puede convertirse en un obstáculo que merme todo el trabajo de recolección y análisis hecho por el investigador, la gestión ética de los datos.

Por sí sola la gestión ética de datos podría dar para escribir una tesis completa o por lo menos un artículo sobre la gestión y tratamiento de bases de datos en la sociología, no obstante, para fines prácticos la entenderemos como una serie de lineamientos utilizados para la administración y uso de datos sensibles<sup>37</sup> de manera responsable,<sup>38</sup> con la finalidad de proteger la privacidad de los datos usados en el análisis durante todo su ciclo de vida hasta su eliminación. Puesto que, al tener un mayor flujo de información circulando, es posible perder de vista la manera en que los datos están siendo procesados, por lo que es necesario tener presente una gestión ética que se encuentre acorde con los propósitos de nuestra investigación.

Ahora bien, sabiendo que el investigador, analista o sociólogo depende la mayor parte del tiempo de bases de datos que le ayuden a estimar el comportamiento social de los individuos y conocer las probabilidades en la evolución de un problema, puede ser posible que incluir una gestión ética de datos sea un impedimento. Pero, por contrario que pueda parecer esa misma gestión ética dentro de Big Data, permitiría tener una mayor objetividad al recolectar información, es decir, siguiendo este criterio, a medida que recolectamos y concentramos los datos en un clúster, estos se van anonimizando, de tal forma que los paquetes de datos no se encuentren referenciados directamente a un individuo; despersonalizando en la medida de lo posible al actor, pero dejando rastros que permitan vincularlo a nuestra muestra.

Por último, respecto a la integración del Big Data como parte de los recursos de análisis y recolección de datos de la sociología, debemos entender que, pese a que los grandes datos han encontrado un elemento vinculante con otras áreas desde hace ya algunos años, hasta la fecha siguen siendo un campo relativamente nuevo para la sociología, por lo que los pocos abordajes vistos sobre el tema han sido teórico-reflexivos con una perspectiva más general a su posible aplicación en las

---

<sup>37</sup> Como datos sensibles, se refiere a datos que se encuentren ligados directamente a los sujetos de estudio. Por ejemplo: nombres, apellidos, direcciones, información bancaria, contraseñas, es decir, cualquier dato que pueda infringir en los códigos o normas relacionadas con la privacidad de los datos.

<sup>38</sup> Esta responsabilidad deriva de manera independiente a las dimensiones de control mostradas en el capítulo uno.

ciencias sociales, pero que sugieren plantear nuevos objetos de estudio relacionados a la digitalización, dejando clara la posibilidad de poder vincular a la sociología con Big Data mediante la adopción de técnicas y metodologías que en principio son atípicas.

<b>Tabla 4. Ventajas y desventajas para hacer análisis de datos a través de Big Data desde una perspectiva sociológica</b>		
<b>Tópico</b>	<b>Ventajas/utilidad</b>	<b>Desventajas/obstáculo</b>
Mayor flujo de información	<ul style="list-style-type: none"> <li>• Un mayor caudal de datos permite ampliar el universo muestral.</li> <li>• Optimiza considerablemente la sistematización de información</li> <li>• Aumento en la calidad de los datos.</li> <li>• Resultados más precisos a corto, mediano o largo plazo.</li> <li>• Mejora en la personalización de los parámetros muestrales en la obtención de datos</li> </ul>	<ul style="list-style-type: none"> <li>• Existe una mayor probabilidad de perder el control de la muestra por el aumento en el flujo de datos sistematizados.</li> <li>• Creación de sesgos de información involuntarios y falta de confiabilidad en los resultados debido a los sesgos.<sup>39</sup></li> </ul>
Versatilidad	<ul style="list-style-type: none"> <li>• Puede incrementar el universo muestral tanto como el investigador o analista lo desee, por lo que es escalable.</li> <li>• Los clústeres producidos se adecuan al proyecto y parámetros deseados por el investigador, por lo que es flexible.</li> </ul>	<ul style="list-style-type: none"> <li>• Los repositorios de información producidos pueden volverse inestables al ampliar, reducir o seccionarlos.</li> <li>• El Big Data tiende a ser volátil por naturaleza</li> </ul>
Proyección de la sociología en los análisis de Big Data	<ul style="list-style-type: none"> <li>• Mejor interpretación del comportamiento social debido a una mayor cantidad de datos analizados.</li> <li>• Adopción de técnicas de recolección de datos no convencionales para la sociología</li> <li>• Al realizar un análisis de Big Data la sociología brinda un mejor soporte teórico-metodológico respecto a la interpretación del comportamiento social.</li> </ul>	<ul style="list-style-type: none"> <li>• Conocimiento necesario requerido para realizar un análisis de Big Data.</li> </ul>

<sup>39</sup> Estos sesgos pueden disminuir o desaparecer introduciendo variables que permitan mantener el control de la muestra.

	<ul style="list-style-type: none"> <li>• El análisis de Big Data no sustituye las técnicas de recolección y análisis en la sociología, sino son complementarias</li> </ul>	
Aplicación multidisciplinaria con perspectiva sociológica	<ul style="list-style-type: none"> <li>• Es un estímulo para la comunidad sociológica para estudiar áreas de interés poco exploradas por la sociología</li> <li>• Mejorar los procesos de búsqueda, recolección y sistematización de datos.</li> <li>• Mejora en el análisis del comportamiento social, extrapolándolo a ambientes digitales</li> </ul>	<ul style="list-style-type: none"> <li>• Tiempo.</li> <li>• Curva de aprendizaje sujeta en función de la formación inicial del interesado y de sus habilidades.</li> </ul>
Portabilidad y Cloud Computing	<ul style="list-style-type: none"> <li>• Utilización de servicios de Cloud Computing como una forma de aumentar la capacidad de los repositorios de datos limitados por el almacenamiento físico.</li> <li>• Mayor capacidad de procesamiento y análisis de información por el cómputo en la nube</li> <li>• Fácil acceso a la información de forma remota.</li> <li>• Es posible trabajar desde cualquier ordenador conectado a internet.</li> </ul>	<ul style="list-style-type: none"> <li>• Costos tasados en función del proveedor del servicio y de la cantidad de almacenamiento requerido.</li> <li>• Como cualquier tecnología de reciente creación, los costos pueden ser elevados.</li> <li>• Cuentas gratuitas limitadas a solo 25GB de almacenamiento</li> </ul>
Gestión ética en la recolección de datos	<ul style="list-style-type: none"> <li>• El uso de técnicas como la minería de datos proporcionaría mayor eficiencia en la extracción de datos, optimizando el proceso de recolección.</li> <li>• Administración y manejo de datos de manera responsable, salvaguardando la privacidad de los datos durante todo el ciclo de vida.</li> <li>• Una Gestión ética acorde con los propósitos de la investigación brinda una mayor objetividad al recolectar la información de nuestro análisis.</li> </ul>	<p>La falta de una gestión ética en la administración de las bases de datos puede generar que el análisis se vuelva obsoleto</p> <p>El tratamiento y la manipulación de datos sin anonimizar puede generar al investigador o analista de datos conflictos, incluso legales por el mal manejo de información.</p>

Fuente: Elaboración propia.

Nota: este cuadro fue elaborado como un resumen a partir de la información presentada en el apartado de conclusiones.

## Bibliografía

### *Referencias bibliográficas*

Barranco Fragoso, Ricardo, *¿Qué es Big Data?*, IBM, México, 2012, disponible en <https://ibm.co/3rnpcC1>

Barros, Alejandro, *Open Data: nuevo paradigma en el manejo de datos*, Revista Bits de ciencia, Universidad de Chile, Chile, 2011, disponible en <https://bit.ly/3jpnj57>

Bayer, *El uso del Big Data en medicina*, Bayer Farmacéutica, España, 2017, disponible en <https://blog.bayer.es/el-uso-del-big-data-en-medicina/>

Bron, Maximiliano (comp.), *Open Data – miradas y perspectivas de los datos abiertos*, Universidad de la Rioja, Ed. La Rioja, España, 2015, disponible en <https://bit.ly/38cHfDv>

Camargo Vega, Juan José; Camargo Ortega, Jonathan Felipe; Joyanes Aguilar, Luis, *Conociendo Big Data*, Facultad de Ingeniería, Universidad Pedagógica y Tecnológica de Colombia, Colombia, 2015, disponible en <https://bit.ly/2O8wb3q>

Elías, Carlos, *Del Social Data al Big Data: evolución de la comunicación política y las campañas electorales en la Era digital*, España, Universidad de la Laguna, 2016, disponible en <https://bit.ly/2NYkAEa>

Falla Gamboa, Jorge, *Datos y Geodatos: geoinformación a tu alcance*, Universidad para la Cooperación Internacional, 2012, disponible en <https://bit.ly/3c3QLtM>

Ferrer Sapena, Antonia & Sánchez Pérez, Enrique A, *Open data: ¿hacia dónde nos dirigimos?*, Anuario ThinkEPI v.7, pp. 150-156, 2013, disponible en <https://bit.ly/30ik9Xz>

Forte, Miguel Ángel; Pignuoli Ocampo, Sergio; Calise Santiago; Palacios, Matías & Zitello, Matías, *Las TIC como problema de la teoría sociológica: una aproximación conceptual a la comunicación digitalizada desde la teoría general de sistemas sociales autorreferenciales y autopoiéticos*, Entramados y Perspectivas Revista de la carrera de sociología, 2012, disponible en <https://bit.ly/3roPYds>

Gutiérrez Puebla, Javier; García Palomares, Juan Carlos; Salas Olmedo, María Henar, *Big (Geo)Data en Ciencias Sociales: Retos y Oportunidades*, Revista de Estudios Andaluces, España, 2016, disponible en <http://oa.upm.es/48020/>

Hernández Leal, Emilcy J.; Duque Méndez, Nestor D.; Moreno Cadavid, Julián, *Big Data: una exploración de investigaciones, tecnologías y casos de aplicación*, Tecnológicas, Instituto Tecnológico Metropolitano, Medellín, Colombia, 2017, disponible en <https://bit.ly/2NYjXui>

Joyanes Aguilar, Luis, *Big Data: análisis de grandes volúmenes de datos en organizaciones*, Ed. Alfaomega, México, 2013. Disponible en <https://bit.ly/3iqWQVt>

Martínez Martínez, Silvia & Lara Navarro, Pablo, *El Big Data transforma la interpretación de los medios sociales*, El profesional de la información, 2014, disponible en <https://bit.ly/389UVin>

Olaya, Víctor, *Sistemas de información geográfica libre y geodatos libres como elementos de desarrollo*, Cuadernos internacionales de tecnología para el desarrollo humano, tecnologías de la información geográfica, Universidad de Extremadura, España, 2009, disponible en <https://core.ac.uk/download/pdf/41787094.pdf>

Oracle, *¿Qué es big data?*, Oracle, México, s/a, disponible en <https://bit.ly/3qokkLE>

Pfizer, *Big Data en el mundo de la salud*, Pfizer Farmacéutica, España, 2016, disponible en <https://bit.ly/2OrjQXI>

Powerdata, *Big Data: ¿En qué consiste? Su importancia, desafíos y gobernabilidad*, España, disponible en <https://www.powerdata.es/big-data>

Powerdata, *Calidad de datos vs mundo real en la integridad de datos*, España, 2013, disponible en <https://bit.ly/3kRrZRI>

Powerdata, *Data Warehouse: todo lo que necesitas saber sobre el almacenamiento de datos*, España, 2014, disponible en <https://www.powerdata.es/data-warehouse>

## *Referencias complementarias*

Acevedo, Lauro, *Sociografía*, El Vigía, Baja California, México, 2017 disponible en <https://bit.ly/30fvYxR>

ADN40, *estas son las redes sociales con más usuarios en 2019*, redacción ADN40, México, 2019, disponible en <https://bit.ly/3edqfRu>

Alonso, Rodrigo & Sánchez, J.M., *Google reconoce que escucha todo lo que hablas en casa*, ABC tecnología, España, 2019, disponible en <https://bit.ly/3uXaELF>

Arcila Calderón, Carlos; Barbosa Caro, Eduar; Cabezuelo Lorenzo, Francisco, *Técnicas big data; análisis de textos a gran escala para la investigación científica y periodística*, El profesional de la información, 2016, disponible en <https://bit.ly/3uUwuPU>

Aristegui noticias, *Se activó sistema de “bots” a favor de AMLO en conversación sobre el primer informe de gobierno*, Aristegui noticias, México, 2019, disponible en <https://bit.ly/30fwl67>

Barrera, Alba, *Así circula el contenido social*, BAQUIA, España, 2013, disponible en <https://bit.ly/3uWNwga>

BBC mundo, *Con un test de Facebook, Cambridge Analytica obtuvo la información de millones de usuarios*, BBC mundo en Animal político, 2018, disponible en <https://bit.ly/3sTuSEk>

BBC noticias, *5 claves para entender el escándalo de Cambridge Analytica que hizo que Facebook perdiera US\$37.000 millones en un día*, BBC mundo, 2018, disponible en <https://bbc.in/3kLDaeD>

BBC noticias, *La página en la que poder ver cómo Facebook te analiza para enviarte anuncios personalizados*, BBC mundo, 2017, disponible en <https://bbc.in/3c2FDxa>

BBVA, *Big Data: ejemplos reales del uso*, BBVA, México, 2018, disponible en <https://www.bbva.com/es/ejemplos-reales-uso-big-data/>

BBVA, *El enigma del Big Data ¿cómo definirlo?*, BBVA, México, 2017, disponible en <https://www.bbva.com/es/enigma-big-data-definirlo/>

Beltrán Martínez, Beatriz, *Minería de datos*, Facultad de ciencias de la computación, Benemérita Universidad Autónoma de Puebla, 2016, disponible en <https://bit.ly/3c92fMq>

Bezanilla, José Manuel & Miranda, Ma. Amparo, *La socionomía y el pensamiento de Jacobo Levy Moreno: Una revisión teórica*, Revista de psicología GEPU, México, 2012, disponible en <https://dialnet.unirioja.es/revista/15248/V/3>

Bezanilla, José Manuel, *Sociometría: un método de investigación psicosocial*, PEI Editorial, México, 2011, disponible en <https://bit.ly/3uSHj5b>

Boeris, Claudia E, *Aplicación de técnicas de análisis de redes sociales y de co-ocurrencia de palabras en la determinación de frentes de investigación*, Instituto Argentino de Radioastronomía, CCT- La Plata, Argentina, 2013, disponible en <https://bit.ly/3kLCUfF>

Brand Monsalve, Edinson, Gabriel & Gómez, Henry, *Análisis de redes sociales como metodología de investigación. Elementos básicos y aplicación*, Facultad de Ciencias Sociales y Humanas, Universidad de Antioquia, 2006, disponible en <https://bit.ly/3rmCFKi>

Camacho, Antonio, *Tendencia en análisis de datos: la aniquilación del dirty data*, Forbes México, México, 2019, disponible en <https://bit.ly/3rpGPkI>

Claverol, Mercè; Simó Ester; Zaragoza Marisa, *Matemática discreta: teoría de grafos*, Departamento Matemática Aplicada IV EPSEVG, Universidad politécnica de Cataluña, Cataluña, 2012, disponible en <https://bit.ly/3qjJtY2>

Cortés Campos, Rocío Leticia; Zapata González, Alfredo; Menéndez Domínguez, Victor Hugo & Canto Herrera, Pedro José, *El estudio de los hábitos de conexión en redes sociales virtuales, por medio de la minería de datos*, Innovación educativa (México, DF), 2015, disponible en <https://bit.ly/3uZyI5W>

Cuenta, Jesús, *Mapa para desmitificar la transformación digital*, Forbes México, México, 2019, disponible en <https://bit.ly/3c8ApA7>

De la Rosa Troyano, Fernando & Martínez Gasca, Rafael, *Sistemas de inteligencia web basados en redes sociales*, Revista hispana para el análisis de redes sociales, 2007, disponible en <https://revistes.uab.cat/redes/article/view/105>

Departamento de orientación, *El sociograma y su utilidad en los grupos de clase*, Colegio San Agustín los negrales, Guadarrama, Madrid, sin año, disponible en <https://bit.ly/30harF2>

El economista España, *¿Qué sucede en internet cada minuto? De 900,000 accesos a Facebook a 3.5 millones de búsquedas de Google*, el economista, España, 2017, disponible en <https://bit.ly/3kKPZpu>

El País, *Las claves para entender el escándalo político de Facebook y Cambridge Analytica*, El País Uruguay, Uruguay, 2018, disponible en <https://bit.ly/3c0e8nY>

Elorriaga Illera, Angeriñe; Merchan Mota, Iker; Vink Larruskain, Nainara, *El Social Big Data: una oportunidad empresarial y laboral*, Estudios sobre el mensaje periodístico, Ediciones complutense, Madrid, 2018, disponible en <https://bit.ly/38d2rcI>

Excelsior, *México, cuarto lugar a nivel mundial en uso de redes sociales*, Excelsior, México, 2018, disponible en <https://bit.ly/3n30iXe>

Fernández Gallardo, Pablo & Fernández Pérez, *Capítulo 9 El lenguaje de los grafos en "El discreto encanto de la matemática"*, Universidad Autónoma de Madrid, España, 2018, disponible en <https://bit.ly/3roxMjS>

Fernández, Yúbal, *Cómo evitar que Google Maps guarde un historial con todas tus ubicaciones*, Xataka, 2017, disponible en <https://bit.ly/2Py8Ubr>

Forbes, *Google admite que escucha conversaciones con su asistente virtual*, Forbes México, México, 2019, disponible en <https://bit.ly/3sQZBBR>

Frankenfield, Jake, *Social Data*, Investopedia, 2020, disponible en <https://bit.ly/30dfhTP>

Fresneda Lorente, Carolina, *El lado más social del Big Data*, Blogthinkbig.com de Telefónica, 2017, disponible en <https://bit.ly/3n0JhNa>

García Miranda, Jesús, *Introducción a la teoría de grafos*, Departamento de álgebra, Universidad de Granada, España, 2005, disponible en <https://bit.ly/2O49I7E>

García Nocetti & Demetrio Fabián, *Analizar millones de millones de datos y orientar la toma de decisiones, objetivos del big data*, Instituto de Investigaciones Matemáticas Aplicadas en Sistemas, UNAM, México, 2018, disponible en <https://bit.ly/3kM0eKh>

García, Carmen, *Big Data, la tecnología que está cambiando el mundo*, IBM, España, 2014, disponible en <https://ibm.co/3kSajoP>

García, Claudia Ivett, *Legislar el ecosistema digital*, Vértigo Político, México, 2020, disponible en <https://bit.ly/30cNu5M>

Garfias, Francisco, *Los “bots” de AMLO y el dinero en redes*, Excelsior, México, 2019, disponible en <https://bit.ly/3bjydXp>

Google, *Google se enorgullece de que los Países Bajos sean el hogar de uno de nuestros centros de datos*, Google, Eemshaven, Países Bajos, 2016, disponible en <https://www.google.com/about/datacenters/locations/eemshaven/>

Goyzuela Rivera, Samuel Israel, *Big Data Marketing: una aproximación*, PERSPECTIVAS, Universidad Católica Boliviana San Pablo Cochabamba, Bolivia, 2015, disponible en <https://www.redalyc.org/articulo.oa?id=425941232007>

Guevara Maldonado, César Byron, *Reconocimiento de patrones para identificación de usuarios en accesos informáticos* (tesis para obtener el grado de master en investigación en informática), Universidad Complutense de Madrid, España, 2012, disponible en <https://bit.ly/3t5mMIX>

Hanneman, Robert A., *Introducción a los métodos del análisis de redes sociales. Capítulo quinto*, Departamento de sociología de la Universidad de California Riverside, Revista Redes, 2002, disponible en <https://bit.ly/3uQZgRo>

Hasbun, Helen, *Big data y la sociología de la comunicación*, periódico Hoy digital, República Dominicana, 2019, disponible en <https://bit.ly/2MSqZQE>

Hernández Armenta, Mauricio, *México ya tiene mas de 74 millones de personas conectadas a internet*, Forbes México, México, 2019, disponible en <https://bit.ly/3sUS1q2>

IBM, *Análisis de datos*, IBM Knowledge Center, 2019, disponible en <https://ibm.co/38fwlr5>

IBM, *Analítica de datos: un proyecto de generación de valor cómo transformar Big Data en resultados, a través de la analítica*, IBM Institute for Business Value, 2013, disponible en <https://bit.ly/3c5a5Xx>

IIC, *Big Data en medicina: aplicaciones útiles*, Instituto de Ingeniería del Conocimiento (IIC), Universidad Autónoma de Madrid, Madrid, España, 2016, disponible en <https://bit.ly/3uVYXoz>

IIC, *Big Data en salud*, Instituto de Ingeniería del Conocimiento (IIC), Universidad Autónoma de Madrid, Madrid, España, 2016, disponible en <https://bit.ly/3ec2cCh>

IIC, *Big Data, aprovecha los datos para extraer información de valor para tu negocio*, Instituto de Ingeniería del Conocimiento (IIC), España, 2019, disponible en <https://www.iic.uam.es/big-data/>

IMPI, *IMPI en cifras enero-diciembre 2018*, Instituto Mexicano de Propiedad Industrial, México, 2018, disponible en <https://bit.ly/3kNbpSR>

INCAP, *Sistema de información*, Instituto de nutrición de Centro América y Panamá, Guatemala, sin año, disponible en <https://bit.ly/3gk3YDK>

INFOBAE, *Qué es (y qué no es) el Big Data y por qué es útil en la educación*, INFOBAE, 2019, disponible en <https://bit.ly/3qjQWX4>

Kuz, Antonia; Falco, Mariana; Giandini, Roxana, *Análisis de redes sociales: un caso práctico*, Computación y sistemas 20(1), 89-106, disponible en <https://bit.ly/3kPxHn1>

Larrosa, Juan MC, *Capturando redes de Facebook con R*, Blog de análisis de redes sociales, Universidad Nacional del Sur, Argentina, 2015, disponible en <https://bit.ly/38dqMia>

Larrosa, Juan MC, *Visualizando Facebook con Gephi*, Blog de análisis de redes sociales, Universidad Nacional del Sur, Argentina, 2018, Disponible en <https://bit.ly/3rmBlqK>

Lozares, Carlos, *La teoría de redes sociales*, Papers Revista de sociología, Universidad Autónoma de Barcelona, Departamento de sociología, Barcelona, España, 1996, disponible en <https://bit.ly/3kZXYiJ>

Martínez Ruíz, Xicotécatl (coord.), *Infoesfera*, Instituto Politécnico Nacional, México, 2015, disponible en

Martínez, León A, *7 gráficos sobre los usuarios de internet en México en 2018*, El economista, México, 2018, disponible en <https://bit.ly/2OtwHbT>

Mejía Llano, Juan Carlos, *Estadísticas de redes sociales 2020: usuarios de Facebook, Instagram, Youtube, LinkedIn, Twitter, Tiktok y otros*, Juan Carlos Mejía Llano, 2020, disponible en <https://bit.ly/3v0u6af>

Mena Roa, Mónica, *¿Qué sucede en internet en un minuto?*, Statista, La agenda digital para España, España, 2020, disponible en <https://bit.ly/2QEGVHP>

Meneses Rocha, María Elena, *Grandes datos, grandes desafíos para las ciencias sociales*, México, Revista mexicana de sociología, 2018. Disponible en <http://www.scielo.org.mx/pdf/rms/v80n2/0188-2503-rms-80-02-415.pdf>

Open Knowledge Foundation *¿Por qué datos abiertos?*, Open Knowledge Foundation, 2014, disponible en <https://opendatahandbook.org/guide/es/why-open-data/>

Open Knowledge Foundation *¿Qué son los datos abiertos?*, Open Knowledge Foundation, 2014, disponible en <http://opendatahandbook.org/guide/es/what-is-open-data/>

Ortiz, Germán, *Hacia dónde deben ir los medios digitales*, Forbes México, México, 2019, disponible en <https://www.forbes.com.mx/hacia-donde-deben-ir-los-medios-digitales-ii/>

Palomera, Iván, *¿Cómo elaborar data analytics de tus clientes?*, Forbes México, México, 2019, disponible en <https://www.forbes.com.mx/como-elaborar-data-analytics-de-tus-clientes/>

Papí Gálvez, Natalia & García Bonal, Raquel, *La comunicación de la innovación de las agencias de medios. Big data en las webs propias*, Revista Internacional de Comunicación Ámbitos, Sevilla, España, 2016, disponible en <https://bit.ly/30dJGRV>

PECAR, *Análisis reticular*, Peña Complutense para el Análisis de Redes (PECAR), Universidad Complutense de Madrid, Madrid, España, 2014, disponible en <https://bit.ly/3epf5t1>

Peco, Ramón, *No es una leyenda urbana, Google escucha a través de los teléfonos*, La vanguardia, México, 2018, disponible en <https://bit.ly/30diin7>

Peersman, Greet, *Sinopsis: Métodos de recolección y análisis de datos en la evaluación de impacto*, Centro de investigaciones de UNICEF, Florencia, 2014, disponible en <https://bit.ly/2QyCyOm>

Pereyra, Diego, *De la sociografía a la sociología científica. Una revisión del itinerario intelectual de Gino Germani*, V jornadas de sociología de la UNLP, Universidad Nacional de la Plata, Facultad de Humanidades y Ciencias de la Educación, Departamento de sociología, La Plata, 2008, disponible en <https://bit.ly/2Pxhib7>

Pfizer, *Big Data en el mundo de la salud*, Pfizer Farmacéutica, España, 2016, disponible en <https://bit.ly/2OrjQXI>

Pfizer, *Pfizer y Mendelian presentan una plataforma tecnológica para apoyar a los profesionales sanitarios en la investigación y el diagnóstico de las enfermedades raras*, Pfizer Farmacéutica, España, 2017, disponible en <https://bit.ly/30msUzP>

Porras Castaño, Javier, *Analítica de datos, inteligencia artificial y Big Data en la banca*, Openmind BBVA, 2018, disponible en <https://bit.ly/3qhptoJ>

Powerdata, *Calidad de datos. Cómo impulsar tu negocio con los datos*, España, disponible en <https://www.powerdata.es/calidad-de-datos>

Powerdata, *Integración de datos: concepto e importancia en la empresa actual*, España, disponible en <https://www.powerdata.es/integracion-de-datos>

Powerdata, *Desmitificando el Data Governance: qué, cuándo, dónde y por qué*, España, disponible en <https://www.powerdata.es/data-governance>

Powerdata, *Data quality: cómo afecta la calidad del dato en los resultados de tu compañía*, España, 2015, disponible en <https://bit.ly/3qnmA5M>

Powerdata, *10 técnicas de análisis de datos para estadísticas de big data*, España, 2017, disponible en <https://bit.ly/3e6XMwy>

Powerdata, *Data center: definición, tipos y tendencias*, España, 2018, disponible en <https://bit.ly/2PrsdTJ>

Powerdata, *Data Lake vs Big Data, principales diferencias*, Redacción Powerdata, España, 2018, disponible en <https://bit.ly/3ecZnkw>

Powerdata, *Datos en tiempo real: ¿cuál es la diferencia entre dato e información?*, España, 2018, disponible en <https://bit.ly/38cdoLx>

Powerdata, *Garantizar la calidad de datos en sistemas informáticos*, España, 2018, disponible en <https://bit.ly/3ednWOM>

Powerdata, *Qualitative data análisis, consejos para aplicarlo con éxito*, España, 2018, disponible en <https://bit.ly/3rmHx22>

Powerdata, *Los pilares de la gestión del big data y arquitectura big data*, España, 2019, disponible en <https://bit.ly/3ed7ptq>

Powerdata, *Dark Data, recomendaciones para sacarle el máximo partido*, España, 2019, disponible en <https://bit.ly/3ck0TPI>

Ramos Vidal, Ignacio, *Análisis de redes sociales: una herramienta efectiva para evaluar coaliciones comunitarias*, Revista de salud pública vol.17, 2015, disponible en <http://www.scielo.org.co/pdf/rsap/v17n3/v17n3a02.pdf>

Rayo, Ángel M, *Análisis de datos en Big Data: tipos y fases del análisis*, Bit, España, 2016, disponible en <https://www.bit.es/knowledge-center/analisis-de-datos-en-big-data/>

Regader, Bertrand, *¿Qué es la sociometría? Características, objetivos y métodos*, Psicología y mente.com, disponible en <https://bit.ly/3rxkvWc>

Rochina, Paula, *El análisis de redes sociales mediante la teoría de grafos*, Revista digital INESEM, Instituto Europeo de Estudios Empresariales (INESEM), Granada, 2017, disponible en <https://revistadigital.inesem.es/informatica-y-tics/teoria-grafos/>

Rodríguez, Antonio & Santamarina Cristino, Patricia, *Análisis del uso de las redes sociales en internet, Facebook y Twitter en las Universidades españolas*, Revista de comunicación y tecnologías emergentes, Madrid, España, 2012, disponible en <https://bit.ly/38cXwIK>

Rodríguez, Patricio; Palomino, Norma; Mondaca, Javier, *El uso de datos masivos y sus técnicas analíticas para el diseño e implementación de políticas públicas en Latinoamérica y el caribe*, Banco Interamericano de Desarrollo, 2017, disponible en <https://bit.ly/3edZWdU>

Romero Calvo, Rocío, *El sociograma*, Revista digital “Innovación y experiencias educativas”, Central Sindical Independiente y de Funcionarios (CSFI), Andalucía, España, 2010, disponible en <https://bit.ly/3uV6sfn>

Said Hung, Elías, *Big data: retos desde la perspectiva de las ciencias sociales*, 2016, Artículo web disponible en <https://bit.ly/2MNUOBK>

Sancho Caparrini, Fernando, *Introducción a las redes complejas*, Departamento de ciencias de la computación e inteligencia artificial, Universidad de Sevilla, Sevilla, 2016, disponible en <http://www.cs.us.es/~fsancho/?e=80>

SAS, *Analítica de big data: ¿Qué es y por qué es importante?*, SAS, 2018, disponible en <https://bit.ly/30gf1mW>

SAS, *Analítica y Big Data: un par indisoluble*, Forbes México, México, 2015, disponible en <https://www.forbes.com.mx/analitica-y-big-data-un-par-indisoluble/>

SAS, *Big Data: What it is and why it matters*, SAS, 2020, disponible en <https://bit.ly/3kPErBl>

Scott, John, *Historia del análisis de redes sociales*, Blog de análisis de redes sociales, Universidad Nacional del Sur, Argentina, 2013, disponible en <https://bit.ly/38e5JMG>

Serrano Cobos, Jorge, *Big data y analítica web. Estudiar las corrientes y pescar en un océano de datos*, El profesional de la información, 2017, disponible en <https://bit.ly/3c8X0fl>

Serrano, Eunáte, *El big data, al alcance de las ciencias sociales*, Tecnopolitica, España, 2014, disponible en <https://bit.ly/3gsrpcl>

Sinergia e Inteligencia de Negocio S.L, *Datawarehouse*, España, disponible en <https://bit.ly/3kOt901>

Sociedad Española de Cardiología, *Big Data en salud: De la medicina basada en la evidencia a la medicina generadora de evidencia*, Sociedad Española de Cardiología, España, 2018, disponible en <https://bit.ly/3uZEJdq>

Sociología necesaria, *Análisis de redes sociales y sociología*, 2013, disponible en <https://bit.ly/3kV35jZ>

Steve, Oscar, *Uno de cada dos hogares en México no tiene internet y la mitad de quienes sí tienen cuentan con una conexión deficiente: INEGI*, Xataka México, México, 2019, disponible en <https://bit.ly/30cPhI2>

Suárez Gonzalo, Sara, *Big social data: límites del modelo **notice and choice** para la protección de la privacidad*, El profesional de la información, 2017, disponible en <https://bit.ly/30jF4JK>

Técnicas de investigación educativa G38, *Análisis de datos en métodos estadísticos*, Universidad del sur, Tuxtla Gutiérrez, Chiapas, 2006, disponible en <https://bit.ly/3qi322U>

The Box Populi, *Sociología y social Big Data*, The Box Populi, 2015, disponible en <https://bit.ly/3qokxhU>

The standard CIO, *los 6 pasos que su organización debe seguir para confiar en Big Data*, 2013, disponible en <https://bit.ly/3ebj7Vs>

Urbiola, Pablo, *¿Qué son los ecosistemas digitales?*, BBVA Research, México, 2018, disponible en <https://www.bbva.com/es/que-son-los-ecosistemas-digitales/>

Velázquez, Alejandro; Marín, Luís Rey, *El valor agregado de las redes sociales: propuesta metodológica para el análisis del capital social*, Revista hispana para el análisis de redes sociales, 2007, disponible en <https://bit.ly/3v6uvHk>

Wiener-Bonner, Danielle, *¿Qué es Cambridge Analytica? Guía para entender el polémico caso del que todo el mundo habla*, CNN noticias, 2018, disponible en <https://cnn.it/3ee3cGa>

World Economic Forum, Switzerland, *Big Data, Big Impact: New Possibilities for International Development*, 2012, disponible en <https://bit.ly/3c73eg3>

Xataka, *¿sabemos cuánto ocupa todo el contenido que hay en internet en este momento?*, Xataka, Webedia Brand Services, España, 2016, disponible en <https://bit.ly/3qfU5qB>

Zaforas, Manuel, *¿Qué puede aportar el Big Data al mundo de la Medicina?*, Paradigma digital, Madrid, España, 2016, disponible en <https://bit.ly/3blGDNY>