La Oferta y Disponibilidad de Contenido Audiovisual en la Era de los Datos Masivos

Carlos Castillo

Universitat Pompeu Fabra

Diciembre 2018

Resumen Ejecutivo

La disponibilidad de nuevos servicios audiovisuales de Vídeo Bajo Demanda (VBD) y Plataformas para Compartir Vídeos (PCV), además de funcionalidades de vídeo en plataformas de Redes Sociales (RRSS), han permitido la realización de algo que hasta hace pocos años resultaba impensable, como es el acceso a gigantescos catálogos de contenido audiovisual de forma legítima e instantánea desde cualquier dispositivo conectado a Internet. El acceso a estos contenidos es mediado por técnicas de búsqueda, recomendación y personalización que utilizan como materia prima datos masivos provenientes del comportamiento de los usuarios. Estas técnicas incluyen la predicción de la relevancia que un contenido pueda tener para un usuario concreto en un contexto determinado, predicción que es llevada a cabo por algoritmos cada vez más sofisticados.

El proceso de selección y priorización de contenidos audiovisuales que era de naturaleza editorial en los medios tradicionales ha sido reemplazado por un proceso automatizado con efectos sobre la disponibilidad y el acceso a dichos contenidos. Estos efectos comportan riesgos incluyendo amenazas a la autonomía individual, a la diversidad del contenido y a colectivos vulnerables.

La autonomía individual se ve amenazada tanto por la asimetría de información que existe entre plataformas y usuarios como por la posibilidad latente de ser manipulados de forma encubierta. La diversidad del contenido se ve amenazada por los riesgos de sobre-personalización y de sesgos en la personalización que cambien el balance de los tipos de contenido que se ofrece a los usuarios. Una serie de colectivos vulnerables también pueden ser afectados por el uso de algoritmos que, basándose en sus patrones de consumo, les sobreexpongan a contenido que sea perjudicial para ellos.

Es posible monitorizar estos riesgos mediante una metodología de observación que mida los efectos de distintas tecnologías de personalización de manera sistemática. La metodología de observación propuesta incluye la creación de perfiles artificiales de usuario, la observación de perfiles naturales a través de informadores y la realización de encuestas para medir los distintos factores de riesgo descritos. Esta metodología puede aplicarse tanto a plataformas de VBD como a PCV y RRSS, en estos últimos dos casos, con algunas modificaciones.

En base a este estudio, se recomienda monitorizar los efectos del uso de datos masivos en la disponibilidad y acceso a contenidos audiovisuales; educar respecto a derechos en materia de autonomía y diversidad y posicionarse a favor de la transparencia; identificar los mecanismos para protección de menores usados en servicios audiovisuales; y colaborar con otras entidades respecto a contenido audiovisual ilegal.

Índice

1 Introducción 1
1.1 La era de los datos masivos 1
1.2 Datos masivos en el sector de servicios 2
1.3 Datos masivos y oferta de contenidos audiovisuales 3
1.4 Contenidos de este informe 4
2 Uso de datos masivos en las plataformas de VBD 6
2.1 Objetivos de una plataforma de VBD 6
2.2 Estimación de relevancia en una plataforma de VBD 8
2.2.1 Estimación de relevancia mediante inferencia probabilística 8
2.2.2 Técnicas actuales para estimar relevancia a gran escala 10
2.3 Técnicas para conducir a los usuarios a una conversión 12
2.3.1 Captura y preparación de datos 12
2.3.2 Selección y ordenamiento 13
2.3.3 Diversificación y presentación 14
3 Riesgos para la disponibilidad y el acceso a contenidos en los servicios de VBD 16
3.1 Riesgos sobre la autonomía de las personas 17
3.1.1 Asimetría de información 17
3.1.2 Manipulación encubierta 19
3.2 Riesgos sobre la diversidad del contenido 21
3.2.1 Sobre-personalización 21
3.2.2 Personalización sesgada 22
4 Evaluando los efectos de las plataformas de VBD en la disponibilidad y acceso a contenidos 24
4.1 Factores a evaluar 24
4.1.1 Factores de autonomía 24
4.1.2 Factores de diversidad 25
4.2 Características deseables de indicadores y metodologías de observación 25
4.3 Indicadores propuestos 26
4.3.1 Indicadores de autonomía 26
4.3.2 Indicadores de diversidad 28
4.4 Metodología de observación propuesta 31
4.4.1 Captura de datos usando perfiles artificiales 31
4.4.2 Captura de datos usando perfiles naturales 32
4.4.3 Encuestas a usuarios de plataformas de VBD 33

5 Plataformas para compartir vídeos y redes sociales 34
5.1 Riesgos en las plataformas para compartir vídeos y en las de redes sociales 34
5.1.1 Riesgos para la autonomía individual y la diversidad del contenido 35
5.1.2 Riesgos de sobreexposición de colectivos vulnerables a contenido potencialmente dañino 36
5.2 Metodología de observación en plataformas para compartir vídeos 39
5.3 Metodología de observación en redes sociales 39
6 Recomendaciones 40
6.1 Monitorizar los efectos del uso de datos masivos en la disponibilidad y acceso a contenidos audiovisuales 40
6.2 Educar respecto a derechos en materia de autonomía y diversidad y posicionarse a favor de la transparencia 41
6.3 Identificar los mecanismos para protección de menores usados en servicios audiovisuales 42
6.4 Colaborar con otras entidades respecto a contenido audiovisual ilegal 43
7 Agradecimientos 44
8 Bibliografía 44

Listado de Figuras

Figura 1: La era de los datos masivos 1
Figura 2: Esquema de uso de datos masivos 2
Figura 3: Declaración de uso de datos en dos servicios de VBD 4
Figura 4: Objetivos de plataformas de VBD basados en el concepto de “conversión” 6
Figura 5: Dos páginas de inicio de la misma plataforma para distintos usuarios 9
Figura 6: Datos usados para estimar la relevancia de un contenido 12
Figura 7: Resultados de una búsqueda para dos usuarios distintos 15
Figura 8: Pantallas 1, 2, y 3 en un dispositivo como un Smart-TV que permite desplazamiento hacia abajo y hacia la derecha 16
Figura 9: Riesgos en los servicios de VBD respecto a la disponibilidad y acceso a contenidos 17
Figura 10: Explicación para contenidos recomendados 18
Figura 11: Cinco de las seis tragicomedias ofrecidas por este servicio en la primera pantalla son producciones propias 20
Figura 12: Ejemplo de cálculo de un indicador 30
Figura 13: Etapas de la captura de datos usando perfiles artificiales 31
Figura 14: Ejemplo de cálculo de un indicador de riesgo para colectivo vulnerable 38
Figura 15: Mecanismo para protección de menores 42

Listado de Tablas

Tabla 1: Ejemplo simplificado de datos históricos. 8
Tabla 2: Indicadores para cada factor de autonomía 27
Tabla 3: Indicadores para cada factor de diversidad 29
Tabla 4: Indicadores para cada factor de riesgo para un colectivo vulnerable 38

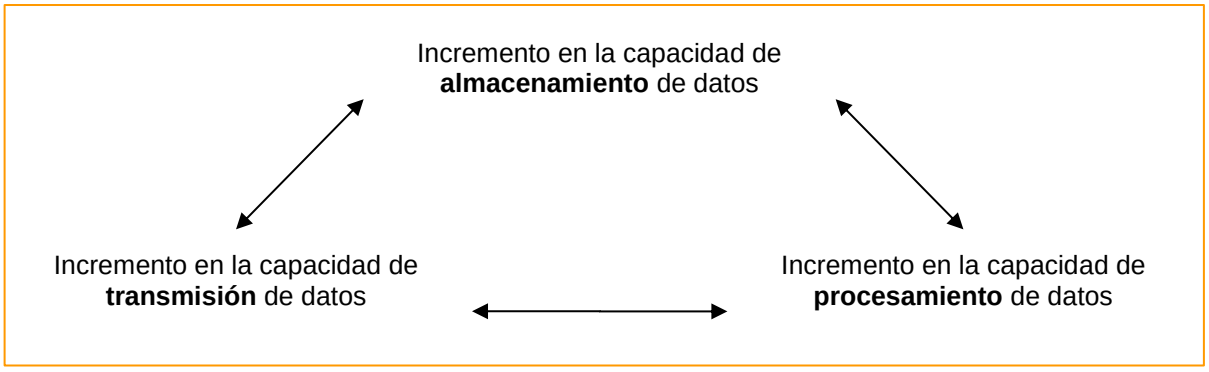
1 Introducción

1.1 La era de los datos masivos

Durante la mayor parte de la historia de la humanidad, la información ha sido almacenada en forma analógica. Inscripciones en piedra, tablillas de arcilla, papiro, papel, e incluso las cintas magnéticas de audio y vídeo, son sustratos análogos en que la información es mapeada directamente a propiedades físicas de un material. Esto significa que la cantidad de información que puede ser almacenada está limitada por barreras inherentes al medio en que se almacena.

El uso de mecanismos digitales para almacenar información, en los cuales números, letras, audio y vídeo son convertidos en código binario (ceros y unos), comenzó con la introducción de los primeros ordenadores digitales en la primera mitad del siglo XX, los cuales llegaron a un público masivo durante los años 1980s y 1990s. Se estima que desde comienzos del siglo XXI la humanidad almacena más información en formato digital que en formato análogo, lo que ha significado un crecimiento de varios órdenes de magnitud en la cantidad de información almacenada y transmitida [Hilbert y López, 2011].

Figura 1: La era de los datos masivos



Esta explosión en la capacidad de almacenamiento de información ha ido de la mano de dos avances tecnológicos complementarios (Figura 1). Primero, el aumento de la capacidad de procesar información, con un importante aumento de la capacidad de cómputo, una caída del costo de los procesadores y una disminución de su tamaño. Segundo, el aumento en la capacidad de transmitir información con un importante aumento de la velocidad de transmisión, una caída del costo por unidad de información y distancia y una migración constante hacia comunicaciones inalámbricas.

Para los ciudadanos, algunos de los elementos más visibles de estos cambios son la ubicuidad del uso de dispositivos móviles de comunicación y cómputo, y el acceso instantáneo a vastas cantidades de información. Sin embargo, hay áreas menos visibles para las personas donde los cambios han sido igualmente profundos. No hay prácticamente ninguna industria que no haya sido transformada radicalmente por el incremento de la capacidad de almacenar, transmitir, y procesar información, resultando en progresivos aumentos de la capacidad productiva y disminuciones en los costos de producción. Los productos que consumimos cada vez más son manufacturados por sistemas robotizados y llegan a nosotros a través de sistemas automatizados de logística, basados en

competencias de adaptación y predicción en tiempo real que utilizan grandes volúmenes de datos, y sistemas ciber-físicos denominados la "Industria 4.0" [BMBF 2016]. Las técnicas que ha desarrollado la industria para integrar estos elementos y trabajar con grandes volúmenes de datos se denominan técnicas de procesamiento de datos masivos, "*Big Data Techniques*," o simplemente "*Big Data*". Siguiendo esta convención, en el resto de este informe cuando hablamos de datos masivos en general nos referimos tanto a los datos en sí mismos como a las técnicas que se usan para procesarlos.

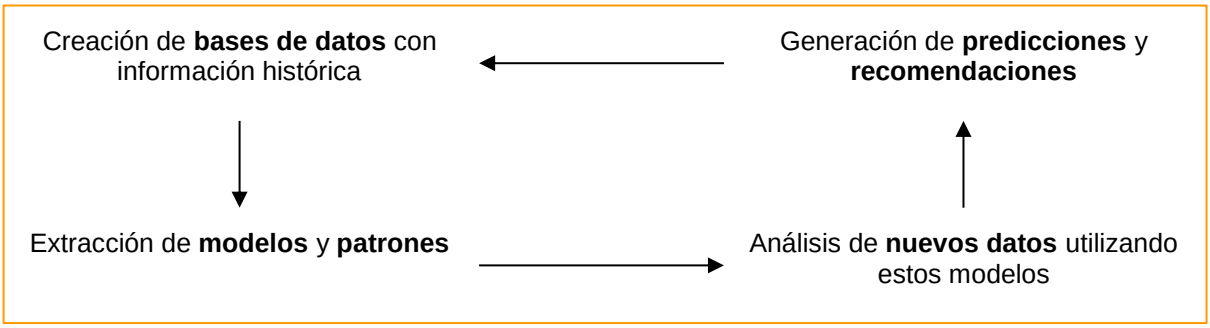
1.2 Datos masivos en el sector de servicios

Al igual que en los sectores primario y secundario, en el sector de servicios hay enormes transformaciones resultantes del aumento en la capacidad de almacenar, transmitir y procesar información. Estas modificaciones también están basadas en datos, pero a diferencia de otros sectores, estos datos en su mayoría no provienen de sensores en la cadena de producción o en información de los proveedores de materias primas o insumos. Estos datos provienen principalmente de los consumidores pasados, presentes y potenciales de estos servicios.

El sector financiero es probablemente uno de los sectores de servicios que tiene la historia más larga respecto al uso de datos masivos. En el mundo anglosajón, es común el uso de un "puntaje de crédito" que refleja con un puntaje positivo el historial de una persona que ha solicitado y pagado créditos en el pasado; en el resto del mundo, es más común el uso de un registro de morosos. En todos los casos, la institución que otorga un crédito a una persona recibe, analiza y decide el riesgo que tiene esta persona de no cumplir con su obligación crediticia, en gran medida basándose en datos que son entregados por el mismo individuo, además de datos que tenga la propia institución. Se trata normalmente de una situación altamente regulada respecto a los datos que pueden ser utilizados en la decisión.

En otros sectores se utilizan básicamente el mismo tipo de técnicas para determinar cuál es el valor anual de un cliente considerando las compras presentes y potenciales del cliente, cuál es la probabilidad de que un cliente actual se cambie a otra empresa (denominado "*churning*"), cuál es la probabilidad de que un cliente anterior vuelva a ser cliente, y así sucesivamente.

Figura 2: Esquema de uso de datos masivos



En prácticamente todos los casos del sector de servicios, el uso de datos masivos ocurre de la misma manera (Figura 2). Primero, se crea una base de datos histórica con el comportamiento de los clientes anteriores que ha tenido una empresa. Segundo, se analiza esa base de datos usando métodos frecuentemente de aprendizaje de máquina estadístico (*"machine learning"*) para obtener patrones, que a veces pueden ser muy simples ("el 70% de las personas que realizan tres o más

llamadas quejándose por el servicio en el mismo mes, dejarán de ser clientes al mes siguiente"), y otras veces mucho más complejos. Tercero, se aplican los patrones aprendidos de la base de datos histórica sobre nuevos elementos, y se genera una predicción. Cuarto, esta predicción es utilizada para tomar una decisión, que puede ser, por ejemplo, ofrecer una promoción o descuento a un cliente para el cual se ha detectado un posible abandono.

Una de las características más evidentes en los sistemas de aprendizaje de máquinas es que mientras mayor sea la cantidad, calidad y diversidad de los datos que se entreguen como entrada más precisa será la predicción o el resultado. Hasta hace pocos años, los rendimientos en términos de precisión obtenida por dato extra eran claramente decrecientes, es decir, aparecían fenómenos de saturación en los cuales pasado un cierto punto más datos no implican más precisión. Actualmente, y con la masificación en la industria de nuevas técnicas analítico/predictivas basadas en redes neuronales, estos rendimientos no son tan decrecientes. Esto significa que el incentivo para utilizar más datos jamás ha sido tan alto como lo es hoy.

En respuesta a estos cambios, el tono de la opinión pública y el marco regulatorio ha ido cambiando. Hay una preocupación evidente por el uso de datos personales, a menudo obtenidos de interacciones con dispositivos digitales que se comunican en forma autónoma, para la creación de perfiles que son usados por empresas para tomar decisiones que eventualmente afectan a personas. Al mismo tiempo, distintas personas ponen distinto énfasis en la protección de sus datos personales, la transparencia en las decisiones tomadas por estas empresas, y el valor que obtienen a través de servicios que son a menudo innovadores y convenientes. El regulador, en el caso europeo a través del Reglamento (UE) 2016/679 General de Protección de Datos, ha buscado proteger el derecho a la privacidad y a la rectificación de datos y decisiones tomadas en base a estos datos, estableciendo límites a la forma en que los datos personales pueden ser capturados, almacenados, utilizados, y estableciendo obligaciones de transparencia respecto a los mismos.

1.3 Datos masivos y oferta de contenidos audiovisuales

La oferta de contenidos audiovisuales en la era de los datos masivos se ha extendido a través de varios servicios de contenido audiovisual a petición conocidos como Vídeo Bajo Demanda (VBD o en inglés VOD) para ordenadores y televisores inteligentes (*Smart TVs*), que han aparecido en los últimos años:

1. Servicios de VBD bajo modalidad de **suscripción** (en inglés: SVOD), donde el consumidor realiza un pago recurrente por acceso a un catálogo, como Netflix, Movistar, y HBO.
2. Servicios de VBD bajo modalidad **transaccional** (en inglés: TVOD), donde el consumidor paga por cada contenido visualizado o descargado, como Amazon Prime o Google Play Movies.
3. Servicios de VBD bajo modalidad de **retransmisión**[1], donde el consumidor tiene acceso en general gratuito a contenidos anteriormente emitidos por un canal de televisión

Adicionalmente, han aparecido nuevas plataformas para compartir vídeos producidos profesionalmente, semi-profesionalmente, o en forma aficionada, como YouTube, Vimeo, o de medios sociales que ofrecen funcionalidades de alojamiento de estos tipos de vídeo, como Facebook e Instagram.

[1] También denominado de "*catch-up*" en inglés.

En cada uno de estos casos, y como todo el resto del sector de servicios, hay captura, transmisión y uso de datos masivos para conseguir ciertos objetivos. El objetivo de negocio principal del proveedor del servicio es incrementar de forma sostenida el número de usuarios y el tiempo que pasan en el servicio. En los casos en que se ofrece contenido patrocinado (publicidad), el proveedor de servicios también tiene incentivos para aumentar la relevancia y el impacto de esta publicidad conociendo mejor los intereses de los usuarios. En los casos en que los usuarios pueden opinar o comentar sobre el contenido, también se busca acrecentar este involucramiento activo como una forma de profundizar y fidelizar la experiencia.

Los datos que conforman estos datos masivos con los cuales cada servicio intenta optimizar sus objetivos provienen, mayoritariamente, de las mismas interacciones de los usuarios con estas plataformas: qué contenidos busca una persona, qué contenidos ve, durante cuánto tiempo, qué día de la semana y a qué hora, y así sucesivamente. Adicionalmente, puede haber datos complementarios provistos de forma voluntaria por los usuarios o como parte de su contrato de servicios. También, y en menor cantidad, pueden hacer uso de datos comprados o intercambiados con terceros, una práctica que es habitual pero cada vez más regulada. En general, los servicios son transparentes respecto a las categorías de datos que utilizan (Figura 3).

Figura 3: Declaración de uso de datos en dos servicios de VBD (Diciembre 2018)[2]

[HBO] "analizará cómo utiliza usted el Servicio, por ejemplo, qué películas y programas de TV ve, su historial de búsquedas, dispositivos utilizados para acceder al Servicio y lugares desde los que se utiliza el Servicio … a partir del uso que haga del Servicio, evaluaremos sus preferencias e intereses en diferentes series y otros contenidos …."	[Netflix] "Obtenemos información sobre ti y sobre tu uso de nuestro servicio, de tus interacciones con nosotros y nuestra publicidad, así como información relacionada con tu ordenador u otro dispositivo utilizado para acceder a nuestro servicio …. para optimizar la selección de contenido, los algoritmos de recomendaciones y la entrega de nuestros servicios;"

1.4 Contenidos de este informe

El contenido de este informe se divide en seis partes.

En la sección 1, enmarcamos, desde una perspectiva histórica, el desarrollo de los datos masivos y cómo estos se desarrollan en el sector de los servicios además de su funcionamiento en la oferta de contenidos.

En la sección 2, explicamos en detalle cuáles son las formas en que los servicios de VBD utilizan metodologías de procesamiento de datos masivos, incluyendo cuál es el objetivo que se persigue, con qué técnicas se persigue este objetivo, y en base a qué datos.

En la sección 3, presentamos las oportunidades y amenazas que puede significar el uso de datos masivos por los servicios de VBD, respecto a su impacto en el acceso y disponibilidad de distintos tipos de contenido audiovisual para los ciudadanos.

[2] HBO: https://es.hboespana.com/data-policy NetFlix: https://help.netflix.com/es/legal/privacy

En la sección 4, describimos un marco de evaluación recomendado para servicios de VBD que permita aproximarse a una medición del impacto descrito más arriba y que, además incorpore elementos cuantitativos.

Las secciones 2, 3, y 4, están centradas en las plataformas de VBD que son el foco de este informe. La sección 5 considera otros servicios en contenido audiovisual que queda a disposición de los usuarios en forma digital, como son las plataformas para compartir vídeos y las redes sociales que permiten compartir contenido audiovisual.

La sección 6 presenta las recomendaciones que se sugieren en base a este estudio.

Necesariamente, algunos aspectos han quedado fuera de este informe, por no ser producto del uso de algoritmos o de la personalización, o por requerir metodologías completamente distintas de observación. Primero, el cumplimiento de la legislación respecto a privacidad y protección de datos en la relación entre las plataformas y los consumidores. Segundo, el cumplimiento de la legislación en materia de protección de menores, incluyendo mecanismos para la verificación de la edad del usuario. Tercero, el cumplimiento de la legislación respecto a contenidos cuya producción o difusión es un delito, por ejemplo, pornografía infantil o tráfico de personas. Cuarto, el cumplimiento de la legislación respecto a categorías publicitarias prohibidas, como la publicidad del tabaco. Quinto, el cumplimiento de la legislación en materia de derecho de autor, incluyendo limitaciones y excepciones. Nos referimos a estos puntos en el apartado de recomendaciones.

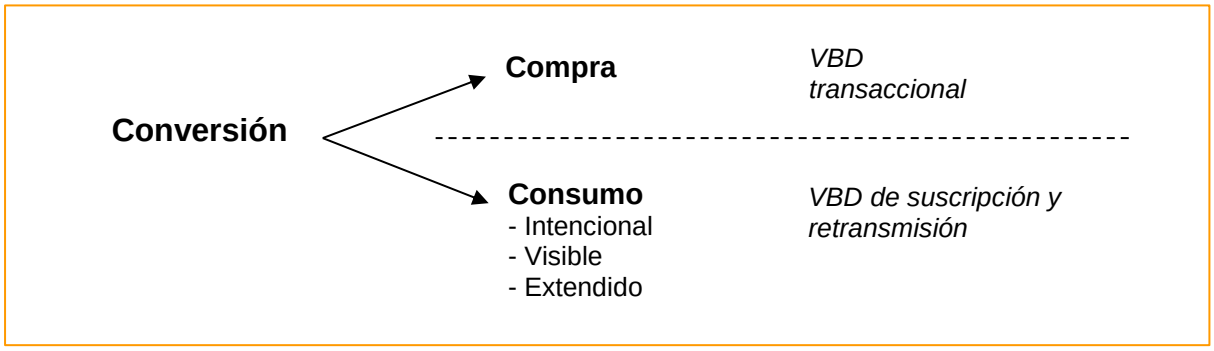
2 Uso de datos masivos en las plataformas de VBD

En un sentido general, las formas en que se usan datos masivos en VBD son similares a las formas en que se usan datos masivos en otras industrias del sector de servicios. El VBD transaccional es el ejemplo más claro ya que se le puede considerar una subcategoría de las plataformas de comercio electrónico. En el VBD de suscripción y retransmisión se observan, también, paralelismos; por ejemplo, en la intención de orientar al consumidor hacia elementos (productos o contenidos) que le resulten relevantes.

2.1 Objetivos de una plataforma de VBD

En el caso del VBD transaccional, el modelo de negocio es conseguir que el consumidor compre un contenido audiovisual. En el VBD de suscripción, también por el tipo de modelo de negocios, es evidente que se pretende retener y expandir la base de suscriptores. En el VBD de retransmisión, los objetivos pueden ser más diversos, pero también consideraremos que retener y aumentar el número de visitantes es probablemente la mayor prioridad.

Figura 4: Objetivos de plataformas de VBD basados en el concepto de "conversión"



En el comercio electrónico y el marketing online, se habla en general de una *conversión* para agrupar distintas acciones deseables que un usuario pueda realizar. Esto ocurre cuando se cumple una meta (táctica) relacionada con un objetivo (estratégico) de negocio. En general, diremos que el diseño y las funcionalidades de una plataforma de VBD tienen como objetivo orientar al usuario hacia una *conversión*. Originalmente el término *conversión* se refiere a un usuario que se convierte en un cliente, pero el término se usa con un significado más amplio para representar cualquier acción deseable para la plataforma que un usuario de Internet ejecuta (Figura 4).

En el caso del VBD de suscripción, la conversión más evidente es la compra de un contenido, pero cabe destacar que también existen otros tipos de conversión que puedan llegar a convertirse en ventas para la plataforma. Un ejemplo sería suscribirse a mensajes o correos electrónicos respecto a una línea de contenido o, instalar una aplicación mediante la cual se puedan enviar ofertas al usuario.

En el caso del VBD de suscripción y retransmisión, la conversión incluye los mismos eventos que en el VBD de suscripción cuando son posibles (ej.: venta de contenido *premium*), y más comúnmente, la

visualización de un contenido audiovisual incluido dentro de la suscripción. Normalmente, para que una visualización pueda ser contada como una conversión debe tener ciertas características:[3]

- **Intencional**, es decir, mediante una acción explícita del usuario y no porque un vídeo sea automáticamente reproducido.
- **Visible** en pantalla, en el caso de que se permita desplazamiento o *scroll* de la pantalla un cierto porcentaje (por ejemplo, el 50%) del video debe ser visible.
- **Extendida** durante un mínimo de tiempo, que puede ser medido en segundos (por ejemplo, 30 segundos) o como porcentaje de la duración del vídeo (por ejemplo, al menos la mitad del vídeo).

Para que las conversiones que involucren un vídeo se produzcan, es fundamental que el contenido sea considerado relevante por el usuario. La **relevancia** en la recuperación de información (ver, por ejemplo, Baeza-Yates y Ribeiro-Neto [2011]), es un concepto difuso que considera varios aspectos: el usuario, el contenido, la forma de presentación y varios elementos de contexto incluyendo tiempo, lugar, estado de ánimo, entre otros. Lo que es relevante para una persona en un lugar y momento puede no ser relevante para la misma en otro lugar y momento.

La relevancia del contenido es una condición necesaria, pero no suficiente para que se produzca una conversión. En otras palabras, analizar la importancia de un contenido es una cuestión de naturaleza probabilística, en la cual el resultado esperado es una aproximación de la probabilidad de que un contenido sea considerado relevante. Ningún sistema actual, ni posible en un futuro próximo, puede predecir con absoluta certeza si una persona encontrará o no un contenido relevante.

Que exista relevancia es fundamental para que pueda haber una conversión, ya que un contenido poco relevante, por definición, tiene baja probabilidad de llevar a una conversión. Más allá de incrementar las conversiones enseñando contenido relevante, pueden existir otros objetivos comerciales o de negocio que pueden llevar a una plataforma a querer que un usuario convierta respecto a un contenido, aunque ese contenido no sea el más relevante para un usuario:

- Puede tratarse de contenido **patrocinado**, en que la plataforma recibe un ingreso extra si el usuario lo visualiza. Estará sujeto a las características nombradas más arriba.
- Puede tratarse de contenido que la plataforma quiera **promocionar** por alguna razón de negocio: un nuevo programa para el cual se está construyendo audiencia, un programa antiguo que potencialmente puede tener impacto debido a una coyuntura noticiosa, un programa de un socio estratégico al que se quiere canalizar audiencia y así sucesivamente.

La inclusión de algunos objetivos de negocio puede o no ser legítima. Existen límites tanto a la publicidad o contenido patrocinado como a la promoción cruzada de servicios. Por ejemplo, la Comisión Europea obligó a Google a abandonar la práctica anticompetitiva de poner sus propios servicios (por ejemplo, de compra online) por encima de los de su competencia en resultados de búsqueda [Miller y Scott, 2014].

Además, teóricamente puede haber otros motivos, no necesariamente de negocio, para que una plataforma busque una conversión al margen de la relevancia que un contenido tenga, incluyendo la promoción ideológica o posición política, o reducir la difusión de un cierto hecho o noticia.

[3] Ver por ejemplo https://www.pressboardmedia.com/how-are-video-views-counted-infographic/

En las siguientes secciones asumiremos que la plataforma conoce el valor de cada objetivo de negocio que no está relacionado con la relevancia. Por ejemplo, el valor de una conversión en contenido patrocinado o bien, la importancia que tiene, por motivos de negocio, la promoción de alguno de sus otros contenidos. Por lo tanto, nos centraremos en describir los mecanismos que puede utilizar para determinar la probabilidad de relevancia de un contenido, y las técnicas mediante las cuales puede conducir a un usuario a una conversión basándose en esta estimación.

2.2 Estimación de relevancia en una plataforma de VBD

A pesar de que cada día cientos de millones de usuarios interactúan con plataformas online que utilizan métodos de datos masivos, existe un desconocimiento generalizado sobre cómo funcionan estos medios. No se hallan estudios específicos sobre el caso de sistemas de VBD, pero para el caso de Facebook, sólo el 20% de los usuarios encuestados sabían que el contenido que ven cuando entran a Facebook es escogido de forma automática por un algoritmo [Rader y Gray, 2015]. En esta sección, presentamos una versión necesariamente simplificada y esquemática de cómo se realiza la estimación de relevancia en una plataforma de VBD.

2.2.1 Estimación de relevancia mediante inferencia probabilística

La relevancia es una cuestión probabilística y para estimarla podemos utilizar métodos probabilísticos. Tomemos un ejemplo muy simplificado en el cual se ha observado a seis usuarios que han visto distintos contenidos (Tabla 1):

Tabla 1: Ejemplo simplificado de datos históricos.

Usuario		**Contenido**		**Contexto**
ID	**Edad**	**ID**	**Género**	**Día de la semana**
U00000001	[20-29]	P00000001	Acción	Viernes
U00000002	[20-29]	P00000001	Acción	Sábado
U00000003	[20-29]	P00000002	Documental	Martes
U00000004	[40-49]	P00000002	Documental	Lunes
U00000005	[40-49]	P00000002	Documental	Jueves
U00000006	[50-59]	P00000003	Drama	Viernes

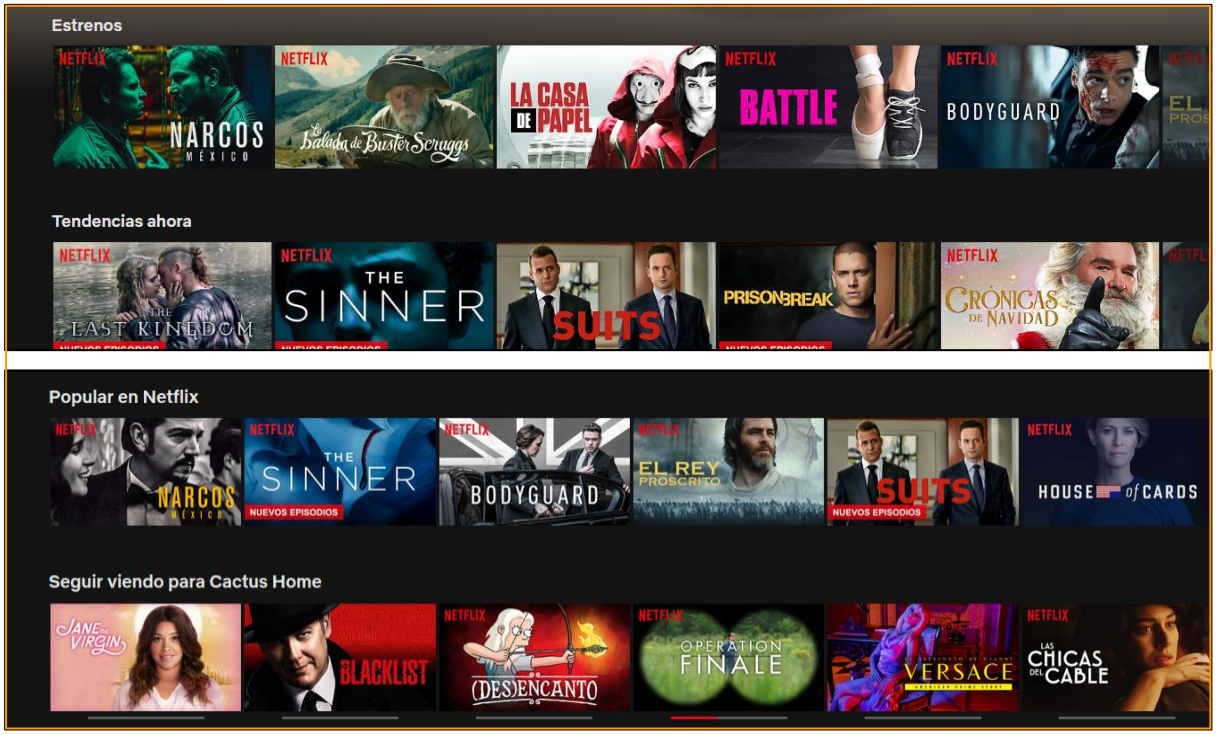
En la tabla hay tres grupos de columnas: los datos del usuario, incluyendo un identificador único, los datos del programa, incluyendo también un identificador único, y los datos del contexto en que ocurrió la conversión.

Si esta fuese toda la información disponible, podríamos llegar a algunas conclusiones que pueden comprobarse mirando la tabla:

- "Los usuarios de 20 a 29 años prefieren los programas de acción a los documentales".
- "Los usuarios de 40 a 59 años prefieren los documentales a otro tipo de programa".
- "Los usuarios de lunes a jueves prefieren documentales, pero viernes y sábado prefieren drama y acción".

Observamos que cada una de estas conclusiones habla de una preferencia expresada en términos tentativos: una inferencia probabilística habla de probabilidades y no de certezas. Esta inferencia puede usarse, por ejemplo, para crear páginas de inicio personalizadas (Figura 5).

Figura 5: Dos páginas de inicio de la misma plataforma para distintos usuarios (páginas obtenidas de una plataforma con sede en España desde dos países europeos)



Estas conclusiones pueden obtenerse mediante la aplicación de algún método de inferencia probabilística que estime la probabilidad de un evento teniendo en cuenta ciertas condiciones. Por ejemplo, en base a estos datos, podríamos estimar la probabilidad de que un programa sea un documental ya que ha sido visto por una persona de entre 40 y 49 años en un día laborable o, más interesante aún, la probabilidad de que una persona de entre 40 y 49 años vea un documental un día laborable. La conexión a través de ambas probabilidades se produce por el teorema de Bayes (matemático del siglo XVIII) y el método específico se denomina inferencia Bayesiana.

Si el proveedor de servicio estima que un contenido es altamente relevante para una persona en un contexto determinado, la probabilidad de que esa persona consuma el contenido será mayor; si por el contrario estima que el contenido no es relevante, la probabilidad será menor. Por lo tanto, una tarea fundamental para la plataforma de VBD es mantener en cada momento y con la mayor exactitud la mejor estimación de relevancia que sea posible.

Desde un punto de vista de procesamiento computacional, el problema al que se enfrenta la plataforma de VBD para realizar este cálculo es la enorme cantidad de observaciones (filas en la

tabla), la enorme cantidad de atributos que puede tener cada usuario y cada programa (columnas en la tabla) y aún más problemático, la cantidad exponencial de combinaciones que hay entre filas y columnas que hacen que sea imposible en la práctica probar todas las combinaciones posibles.

Evidentemente, la solución es probar únicamente algunas de las combinaciones posibles realizando una **búsqueda sistemática**, idealmente con garantías de que las conclusiones obtenidas sean certeras y generales. Por la cantidad de datos que deben ser procesados, esta búsqueda es realizada por un ordenador que obedece ciertas instrucciones descritas en un lenguaje computacional mediante un programa, que es la implementación en lenguaje de máquina de un algoritmo.

Un **algoritmo** es una descripción abstracta y sin ambigüedades de un procedimiento, mediante una serie de pasos y condiciones. El nombre "algoritmo" deriva de la latinización en el siglo XII del nombre del matemático persa Al-Khwārizmī a "*algorismus*". Estamos rodeados de algoritmos, pero normalmente no les llamamos así. Por ejemplo, el método para sumar números de varias cifras usando resto que se enseña en primaria, es una descripción:

- **abstracta**, porque no depende de los números específicos sino que es general,
- **sin ambigüedades**, porque una vez que uno conoce el algoritmo, no necesita hacer preguntas para ejecutarlo,
- **basada en pasos**, porque se calcula una columna de números a la vez,
- **basada en condiciones,** por ejemplo, si la suma excede 10, pasar el resto a la columna siguiente.

En educación superior también se enseñan algoritmos, estos más complejos, como los usados para determinar configuraciones orbitales de distintos elementos químicos o para resolver un sistema de ecuaciones lineales.

Los algoritmos utilizados para realizar inferencia probabilística son complejos, pero esencialmente tienen la misma naturaleza que los algoritmos clásicos que se enseñan en primaria y secundaria: observar una serie de elementos de datos en un orden determinado, realizar un número de operaciones sobre estos datos y detenerse cuando se cumplen ciertas condiciones.

2.2.2 Técnicas actuales para estimar relevancia a gran escala

Del ejemplo que hemos visto de cómo extraer dependencias probabilísticas para estimar la relevancia podríamos plantear algunas preguntas:

- ¿Qué debemos hacer respecto a los eventos no observados? En la tabla vemos que un usuario vio un programa determinado, pero lo que no sabemos es qué otros tipos de contenidos tuvo oportunidad de ver pero decidió no hacerlo. Esto también puede ser muy importante.
- ¿Deberíamos contar cada conversión igual? Algunos de estos contenidos son más de nicho y otros van dirigidos a un público más general. Ver un contenido de nicho puede ser una señal más clara sobre los intereses de un usuario.
- ¿Es cada visualización una señal de interés? Por omisión que un usuario vea un contenido de comienzo a fin puede ser interpretado como una señal de interés, pero también se puede preguntar explícitamente al usuario si el programa que ha visto le ha gustado o no, o permitirle expresar esta opinión tras ver un programa, y algunas plataformas lo hacen.

- ¿Cuántas observaciones son necesarias para llegar a una conclusión? Si nos basamos exclusivamente en la tabla, diríamos que las personas de 40 a 49 años solo quieren ver documentales lo cual probablemente no sea cierto. Ello es el resultado de la falta de observaciones sobre este grupo.
- ¿Qué hacer con un usuario que se inicia en la plataforma? Esto se denomina el problema de "arranque en frío" (*cold start*) y se refiere a la incertidumbre aumentada que existe cuando se intenta determinar la relevancia de un contenido para un usuario nuevo, del cual aún no se ha observado un historial de interacciones.
- ¿Qué tipo de conclusiones son interesantes? Podríamos priorizar conclusiones certeras que se apliquen cuando se cumplen muchas condiciones (edad y contexto, por ejemplo) o conclusiones más generales que se aplican a grupos de usuarios más grandes.

Estas preguntas no agotan la complejidad que existe detrás de la producción de un método de aprendizaje de máquina estadístico efectivo (certero) y eficiente (que no requiera una cantidad exagerada de tiempo o capacidad de cómputo). De hecho, existe un área muy activa de investigación en inteligencia artificial y aprendizaje de máquina en la que trabajan miles de investigadores, y donde se invierten miles de millones de euros cada año.

Algunos elementos tecnológicos que se repiten en esta área en general, y en particular cuando se habla de plataformas de VBD, son métodos para la reducción de dimensionalidad y métodos de *multi-armed bandits* (apodados métodos de "tragaperras").

Los métodos de **reducción de dimensionalidad** intentan reducir un gran número de atributos a una pequeña lista de características implícitas o latentes de una persona o de un programa (ver, por ejemplo, Bell y Koren [2007]). Estas características en algunos casos pueden ser interpretadas como aspectos generales demográficos o de personalidad de un usuario, aspectos generales de un grupo de películas (como por ejemplo "escapistas" versus "realistas"), o aspectos generales de un contexto ("en casa"). Una vez realizado este proceso, se puede elaborar una recomendación con un menor uso de recursos computacionales al reducirse el número de atributos sobre los que se debe operar. También hay mejoras en la certeza de las predicciones al reducirse el nivel de "ruido" en los datos.

Los métodos de ***multi-armed bandits*** (ejemplo, Li et al. [2010]), adoptan un paradigma probabilístico que puede explicarse con la siguiente analogía. Supongamos que nos encontramos delante de un número de máquinas tragaperras, cada una de las cuales tiene una cierta probabilidad de dar un premio. Si desconocemos la probabilidad de dar un premio de cada máquina tenemos que *explorar,* lo que quiere decir ir probando con distintas máquinas. Una vez que encontramos una máquina que parece premiar con una alta probabilidad la tenemos que *explotar*, lo que significa seguir apostando en esa máquina. En todo momento, hay un equilibro entre explorar y explotar. Lo mismo ocurre en una plataforma de VBD que quiere, al mismo tiempo que nos entrega contenido que nos resulte relevante ("explotar" lo que saben de nosotros), aprender qué otro tipo de contenido que aún no nos ha enseñado puede sernos relevante también ("explorar" otras posibilidades). Un algoritmo de *multi-armed bandit* debe, por lo tanto, determinar cuál es este equilibro en forma continuada en el tiempo, explorando para conocer más de los usuarios y enseñándoles contenido relevante para maximizar la posibilidad de conversión en forma sostenida.

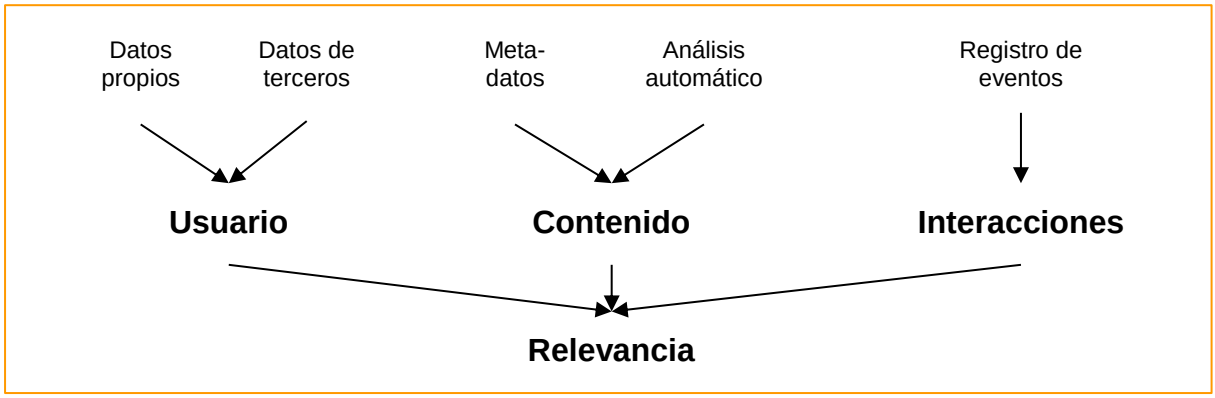
2.3 Técnicas para conducir a los usuarios a una conversión

Cada plataforma de VBD utiliza técnicas propias protegidas por secretos de negocios y patentes para maximizar la probabilidad de una conversión valiosa de sus usuarios. Lo que se conoce sobre el funcionamiento de estas plataformas proviene de sus patentes, de lo que revelan a través de charlas técnicas y artículos científico/técnicos, y de lo que puede ser observado directamente por los usuarios de estas plataformas. El ingrediente principal siempre es la capacidad de estimar la relevancia que tiene para un usuario concreto un programa determinado en un contexto dado.

2.3.1 Captura y preparación de datos

La mayor parte de los datos utilizados por una plataforma de VBD (Figura 6) provienen de las **interacciones** de los mismos usuarios con la plataforma. De hecho, el mismo VP de Innovación de Netflix ha indicado que la interacción es más importante que aquello que los usuarios declaran explícitamente mediante ratings: "lo que quieres y lo que dices que quieres es muy diferente" [Lowensohn 2015].

Figura 6: Datos usados para estimar la relevancia de un contenido



Los datos de interacción son capturados automáticamente. Cada vez que un usuario interactúa con un elemento de la plataforma, incluso en casos relativamente insignificantes como por ejemplo, un botón para pausar un vídeo, se graba en un registro de eventos de la plataforma la fecha, hora y el evento observado (usuario X ha pausado vídeo Y en el minuto Z), además de la información del dispositivo que el usuario estaba usando, y otros datos relevantes del contexto. El proceso mediante el cual se implementa en una plataforma esta captura de datos se llama instrumentación (*instrumentation*).

Los registros de eventos son procesados de forma automática para producir una tabla de datos similar en esencia a la que presentamos anteriormente, pero con un nivel de detalle mucho más fino. Netflix, por ejemplo, utiliza variables como las películas que un usuario ha visto, el género de las mismas, las interacciones con la película, el país del usuario, el idioma que tiene en sus preferencias, el dispositivo usado, la dirección de Internet (IP) usada, la hora del día, y el día de la semana [Chandrashekar et al. 2017], y permite a los usuarios descargar estos datos [Netflix Personal Information Page 2018]. HBO utiliza datos similares, indicando que la captura de estos datos es necesaria para ofrecer el servicio: "usted no puede oponerse a que tratemos la totalidad de sus datos personales y siga exigiéndonos que prestemos el Servicio" [HBO Política de Datos Personales 2018]. Debemos notar que la RGPD en su considerando 32 establece que el consentimiento del

usuario al tratamiento de sus datos personales debe ser obtenido para cada fin por separado. Por lo tanto, sólo es legítimo pedir el consentimiento sobre la "totalidad" de los datos personales si todos y cada uno de esos datos son estrictamente necesarios para prestar el servicio.

Otra fuente de datos es la **información del contenido** audiovisual. Esto incluye metadatos en forma legible por un sistema informático provistos por quién licencia el contenido (duración del contenido, género, créditos, presupuesto, contenidos relacionados, y otros elementos similares). También puede incluir elementos obtenidos automáticamente como puede ser la transcripción del discurso hablado que haya dentro del contenido; información obtenida manual o automáticamente sobre el estado de ánimo, la estética, o el ritmo de una película [Chhabra et al. 2017]; o el reconocimiento automático de objetos, escenas y otros elementos visuales que puedan ser convertidos en palabras clave o descripciones a ser usadas para determinar la relevancia para un usuario.

La información de un **usuario** comienza con los mismos datos de registro en la plataforma. La información puede ser complementada con fuentes externas, por ejemplo, cuando un usuario de una plataforma de VBD asocia su cuenta de Facebook a su cuenta de la plataforma, puede haber un intercambio de datos entre ambas siempre que sea con el consentimiento del usuario, que puede ser libre e informado en algunas ocasiones, o que puede adolecer de vicios. De hecho, en el caso de Facebook la forma en que se comunica a los usuarios qué datos son entregados a terceros para fines publicitarios tiende a ser bastante imprecisa [Andreou et al. 2018, Dance et al. 2018].

Basándose en estos datos, pueden realizarse varias operaciones de preprocesamiento de los datos, que incluyen:

- Reducción de dimensionalidad, como se ha explicado anteriormente, encontrando factores latentes de cada usuario y programas basados en toda la información disponible sobre ellos.
- Cómputo de similaridad, el cual busca determinar similitudes entre usuarios, contenidos y contextos. La relevancia puede aproximarse de forma más precisa usando señales de relevancia de usuarios, contenidos, o contextos similares.
- Perfilado en base a atributos explícitos, en que se busca categorizar a un usuario de forma probabilística dentro de una categoría específica e interpretable, como por ejemplo "usuarios que ven deporte en vivo" o "usuarios que ven series durante varias horas los fines de semana."

Las operaciones específicas que realice cada plataforma dependen de las propias tecnologías que utilice y haya desarrollado cada plataforma, y como hemos dicho antes, son en parte secretos de negocio de cada plataforma.

2.3.2 Selección y ordenamiento

Uno de los objetivos de las plataformas de VBD es producir una experiencia lo más libre de "fricción" posible, es decir, evitarle al usuario tener que interactuar prolongadamente o tener que realizar un esfuerzo cognitivo significativo antes de comenzar a disfrutar del contenido. Por este motivo, y basándose en una estimación de la probabilidad de que ciertos contenidos sean relevantes para un usuario en el contexto actual, se realiza un proceso de selección y orden de los contenidos.

Normalmente, los contenidos serán desplegados en una página del proveedor de VBD de tal forma que **las posiciones más visibles sean ocupadas por los contenidos sobre los cuales se desea o**

espera una conversión y las posiciones menos visibles, por los contenidos sobre los cuales esta conversión es menos deseable o esperable. *Deseable*, en este ámbito, se refiere a objetivos de negocio y *esperable*, respecto a la probabilidad estimada de que el contenido sea relevante. La *visibilidad* de distintas posiciones en una página puede obtenerse mediante experimentos de *eye tracking* en laboratorio y también, mediante registros de clics realizados por los usuarios en un sistema.

Existen al menos tres ámbitos dentro de la plataforma de VBD en los que ocurre esta operación. El primero y el más importante es la **página inicial** de bienvenida, en la cual se recibe al usuario; esta página es la más relevante dado que es la que normalmente verán los usuarios. El segundo ámbito son las **páginas por categoría** dedicadas a distintos géneros y tipologías de contenido, donde los contenidos son filtrados según el género y la categoría.

El tercer ámbito es la **búsqueda por palabras clave**, una funcionalidad que es frecuentemente ofrecida en VBD. En este caso, hay un tercer elemento que se suma a la estimación de relevancia y a los objetivos de negocio: la relevancia de un contenido respecto a palabras clave. Para determinar este último elemento, se utilizan técnicas de recuperación de información multimedia [Lew et al., 2006] similares a las usadas en los buscadores de páginas web. Las búsquedas son también un elemento de interacción y son almacenadas en el registro de eventos asociado a un usuario, para mejorar el cómputo de relevancia en futuras búsquedas.

2.3.3 Diversificación y presentación

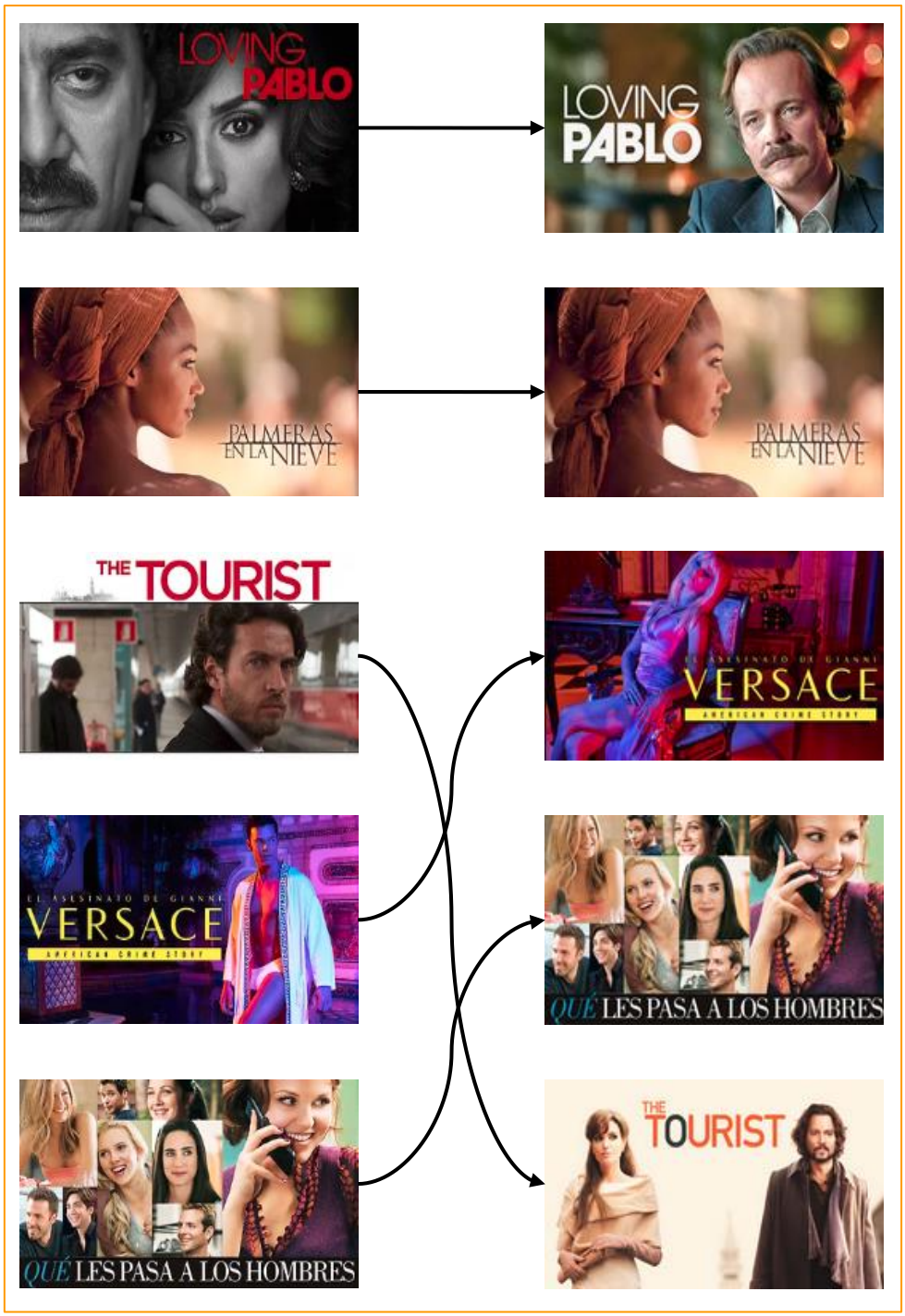
Hay un aspecto adicional que la plataforma de VBD tiene en cuenta para realizar una recomendación y es que los elementos en una pantalla interactúan entre sí y pueden complementarse o competir por la atención del usuario. Un contenido en una determinada posición tendrá una probabilidad distinta de generar una conversión dependiendo de los contenidos que se desplieguen en torno a él. En ese sentido, y aunque por simplicidad computacional una plataforma pueda optar por realizar un cómputo separado para cada posición, en algún momento es necesario considerar estos fenómenos de complementariedad y competencia.

La operación más frecuente para conseguir este objetivo es la **diversificación**, que significa asegurarse de que una página contenga contenidos que sean diferentes entre sí. Esto se realiza por dos motivos, primero porque un contenido puede no tener relevancia para un usuario si otro contenido ya está desplegado (por ejemplo, si el usuario está decidiendo qué serie ver, basta con enseñar uno o dos capítulos de cada serie en la interfaz) y segundo, porque siempre habrá una cierta ambigüedad en las necesidades del usuario. Por ejemplo, si un usuario busca por el nombre de un actor o actriz de larga trayectoria, la plataforma no sabe si busca los trabajos más recientes o los trabajos más clásicos, y debería mostrar una mezcla de ambos tipos de contenidos para asegurarse de que alguno de ellos cumple con las expectativas de los usuarios.

Finalmente, la atractividad de distintos elementos en una página puede mejorarse considerando apropiadamente la forma en que cada contenido se despliega (Figura 7), la cual puede ser personalizada en diferentes carátulas o escenas destacadas de un contenido para cada usuario [Chandrashekar et al. 2017].

Figura 7: Resultados de una búsqueda para dos usuarios distintos
(La búsqueda es "Penélope Cruz" y se trata de dos usuarios del mismo servicio, con sede en España, en la misma fecha, noviembre 2018, pero en países distintos de Europa. Tanto el orden como la presentación de algunos contenidos son diferentes; nótese, por ejemplo, la serie "American Crime Story / Versace")

.

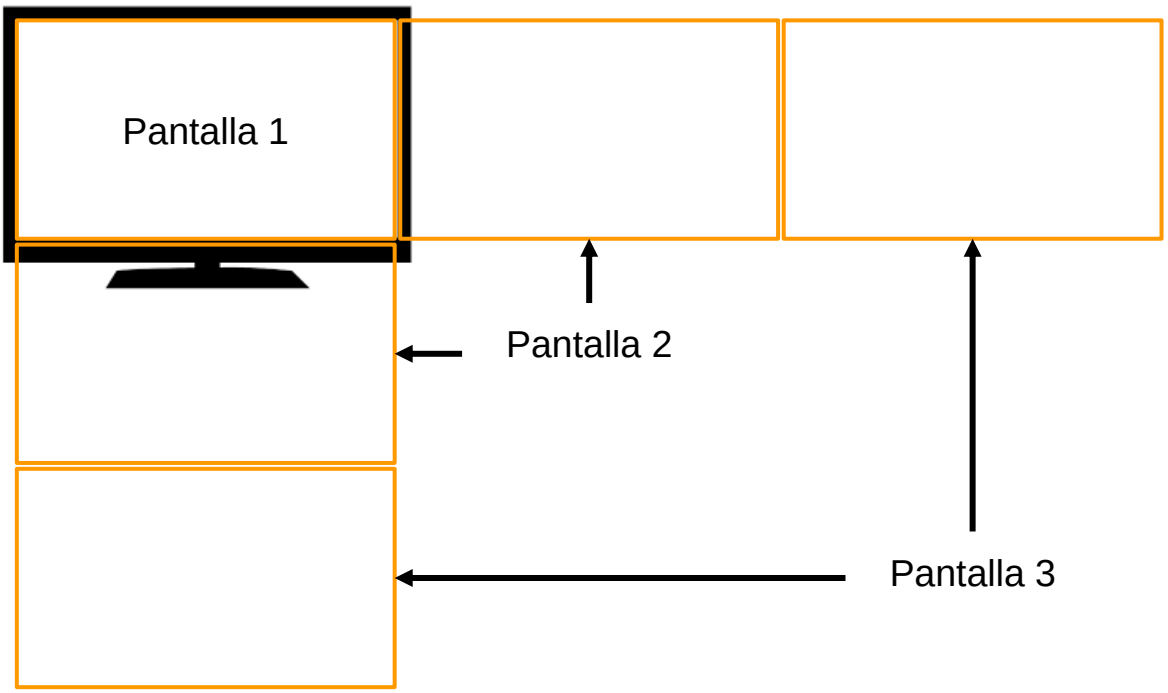


3 Riesgos para la disponibilidad y el acceso a contenidos en los servicios de VBD

A pesar de que el uso de VBD podría no estar cambiando el volumen de contenido audiovisual consumido por el público (en términos de minutos por día, por ejemplo), sí está cambiando la composición del mismo [Broadband TV News 2018]. En esta transición desde contenido seleccionado editorialmente a contenido seleccionado y priorizado automáticamente, es natural que haya consecuencias sobre la disponibilidad y acceso a contenidos de distinto tipo.

La **disponibilidad de contenido** no se refiere solamente al hecho de que un tipo de contenido forme parte del catálogo de una plataforma de VBD, sino a una serie de factores que afectan la probabilidad de que ese contenido sea visionado: la prominencia que se le da en distintos tipos de menú, la frecuencia con que es ofrecido a los usuarios, la saliencia de la forma en que es ofrecido (ej.: destacado), y lo fácil o difícil que resulta localizarlo mediante búsquedas por palabras clave. En otras palabras, no todos los contenidos en el catálogo de una plataforma de VBD están igualmente disponibles, por lo tanto las obligaciones conocidas como "*must-carry*", que indican que un tipo de contenido debe estar disponible, deben convertirse en la obligación de "*must-be-findable*", que el contenido sea encontrable [European Broadcasting Union, 2013]..

Figura 8: Pantallas 1, 2, y 3 en un dispositivo como un Smart-TV que permite desplazamiento hacia abajo y hacia la derecha

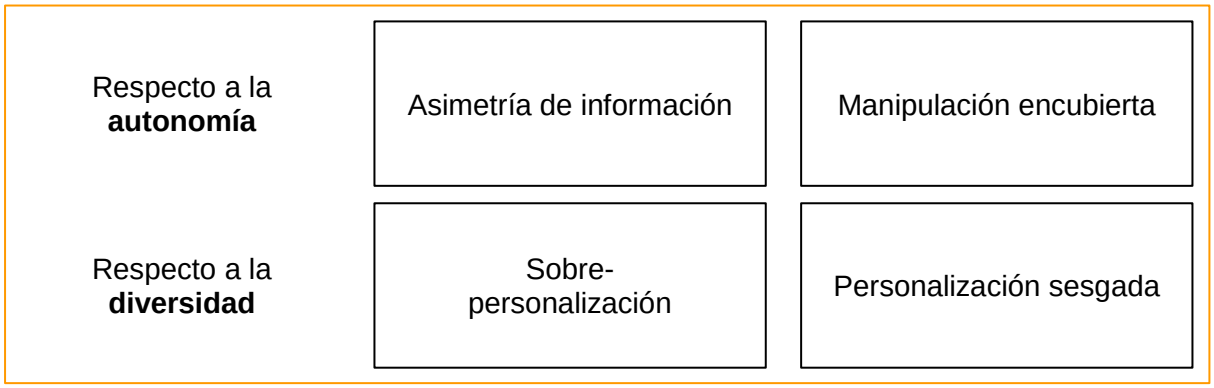


Un elemento muy importante para que un contenido que está presente puede ser considerado encontrable es su posición en un listado, en particular su relación con el contenido que cabe en la primera **pantalla** (Figura 8) de una página en un servicio de VBD [Petit 2018]. Todo lo que es visible sin necesidad de desplazarse (hacer *scrolling*) pertenece a la primera pantalla. Las siguientes pantallas se definen en función de este desplazamiento como segunda pantalla, tercera pantalla, y siguientes. Por ejemplo, la tercera pantalla requiere desplazarse una cantidad equivalente a dos

veces la primera pantalla. Este desplazamiento puede ser en cualquier dirección: normalmente hacia abajo o hacia la derecha; consideramos que la segunda pantalla incluye conjuntamente lo que está abajo y hacia la derecha de la primera pantalla. Evidentemente, hay un punto en el que se debe considerar que un contenido no tiene suficiente prominencia porque no es encontrable, aun cuando esté presente en un listado.

En esta sección, analizamos las posibles consecuencias negativas, algunas de ellas inesperadas, que puede tener el uso de datos masivos en la disponibilidad de contenidos. El foco está puesto en dos valores que pueden ser afectados negativamente: la **autonomía** del individuo y la **diversidad** de los contenidos (Figura 9). Hacemos notar que en esta sección consideramos una definición amplia de contenido que incluye todas sus tipologías (por ejemplo, de ficción, editorial, publicitarios), considerándolos simplemente distintos tipos de oferta en una plataforma de VBD.

Figura 9: Riesgos en los servicios de VBD respecto a la disponibilidad y acceso a contenidos



3.1 Riesgos sobre la autonomía de las personas

La autonomía de las personas es la capacidad de cada individuo de autogobernarse, de tomar decisiones independientes, informadas, y sin coerción. El uso de datos masivos en las plataformas de VBD es una manera de utilizar la tecnología para influir en el comportamiento de las personas [Schwab 2018]. En el caso del VBD el comportamiento sobre el que se influye es esencialmente el consumo de contenido audiovisual, normalmente con fines de entretenimiento, que es un área importante aunque parezca superficialmente trivial.

Existen, al menos, dos maneras en las que el uso de datos masivos en una plataforma de VBD influye sobre la autonomía de las personas: produce una asimetría de información y abre la posibilidad de realizar una manipulación subrepticia.

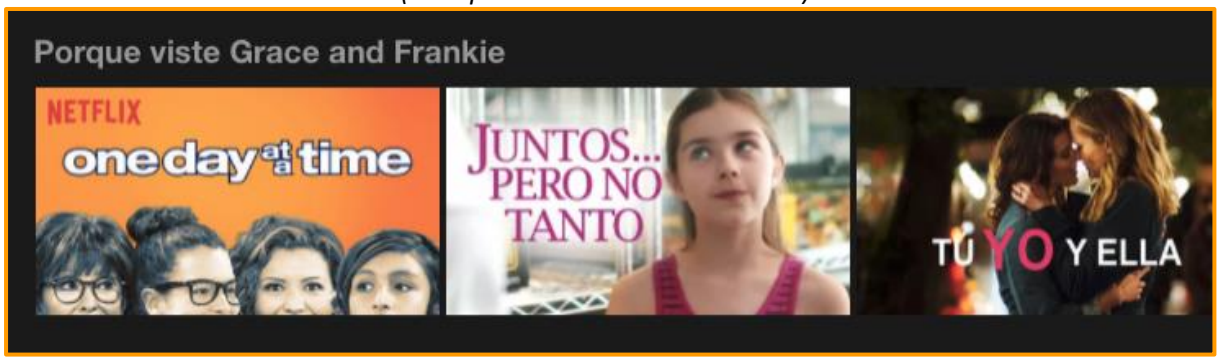
3.1.1 Asimetría de información

Desde los inicios de la televisión abierta con fines comerciales han existido incentivos para que el emisor conozca a la audiencia en detalle e intente influir sistemáticamente sobre su comportamiento. Algunas veces, para incrementar su propia influencia, otras en beneficio de alguna empresa, organización, persona o ideología. Hasta hace pocos años lo que no existía era la capacidad tecnológica para hacerlo con la precisión con la que es posible hoy.

Anteriormente, la audiencia sólo podía ser parcialmente conocida por el *broadcaster* basándose en muestras relativamente pequeñas respecto a la población total y en aproximaciones inexactas. En cambio, hoy se tienen datos mucho más completos acerca de *todos* los que están interactuando con una plataforma y de *todas* sus interacciones. Esto ha resultado un cambio en el balance de poder respecto a la situación anterior, en la cual "las organizaciones de medios sólo podían perseguir pero nunca realmente controlar la atención y el comportamiento de la audiencia" [Arnold 2016]. La relación actual es de profunda asimetría entre la información de la que dispone una plataforma de VBD sobre los usuarios, y la información de la que disponen los usuarios sobre ellas.

Por ejemplo, una plataforma tiene muchísima información relevante para decidir qué contenido recomendar a qué usuarios. En cambio, los usuarios tienen muy poca información relevante para entender y evaluar esa recomendación. De hecho, esta información a menudo no está disponible para los usuarios o es provista de forma muy simplificada en forma de: "Porque viste ... te recomendamos ..." (Figura 10). Esta parsimonia puede ser entendida como un objetivo de diseño, en que se busca como un bien en sí mismo el generar una experiencia en que el algoritmo sea invisible [Hamilton et al. 2014].

Figura 10: Explicación para contenidos recomendados ("Porque viste Grace and Frankie")



La información que se entrega a los usuarios y/o la explicación del funcionamiento de un algoritmo, en los pocos casos en que está presente, puede resultar incompleta e incluso engañosa. Por ejemplo, no está claro que sea *exclusivamente* porque uno ha visto el contenido X que recibe una recomendación, depende de cuándo lo haya hecho, en qué contexto, y en muchos otros de los contenidos que el usuario puede haber consumido. Si bien, hay poca investigación sobre este tipo de explicaciones en las plataformas de VBD, en el caso de los anuncios en Facebook, que también son generados en base al análisis de datos masivos de comportamiento, la explicación que un usuario puede recibir de por qué se le ha enseñado un cierto anuncio ha sido evaluada como "incompleta y engañosa" [Andreou et al. 2018].

Al margen de la forma en que las recomendaciones sean explicadas o no, la interfaz de una plataforma de VBD establece las acciones que son posibles dentro de esta plataforma y la facilidad o dificultad con que esas acciones son realizadas, lo que se conoce como el "ofrecimiento" (en inglés, *affordance*) de un programa computacional interactivo. Dentro de esas posibilidades, es esperable que en un contexto de entretenimiento el usuario tienda a realizar las acciones que resultan más fáciles por la forma en que está diseñada la interacción. Esa acción es, frecuentemente, aceptar alguna de las recomendaciones; de hecho, se estima que entre un 75% [MacKenzie et al. 2013] y un 80% [Chhabra 2017] del contenido consumido en Netflix provenga de recomendaciones.

Esto no significa que los usuarios automáticamente acepten el primer ítem que se les recomienda, de hecho, un usuario típico de VBD puede revisar 40 o 50 títulos antes de escoger uno [Chhabra 2017], pero es innegable que hay una fuerte influencia del sistema de recomendación en lo que el usuario decide consumir finalmente. Incluso un usuario aparentemente selectivo, aquél que revisa muchos títulos antes de escoger uno, está explorando una porción relativamente insignificante del catálogo, limitada a lo que el algoritmo de recomendación le ha ofrecido. La posibilidad de tener tantas opciones quizás "enmascara formas más profundas de manipulación individual" [Arnold 2016] produciendo la ilusión de que se ha tomado una decisión libre y autónoma.

El mayor riesgo para la autonomía de las personas es que se establezca una relación de *dependencia* en la cual un usuario ejerce cada vez menos su propia facultad de decidir, porque sin tener ni la información suficiente ni un diseño de interacción que le permita y aliente a decidir por sí mismo(a), resulta mucho más fácil abdicar estas decisiones en un sistema automatizado.

3.1.2 Manipulación encubierta

Dado que no todo el contenido audiovisual existente cabe en una pantalla, debería resultar evidente para todos los usuarios que, de algún modo, ese contenido ha sido seleccionado. Es decir, que de alguna forma hay una "manipulación" que ocurre con consentimiento mutuo [Koene et al. 2015]. Cabe destacar, sin embargo, que no para todos los usuarios está claro que esta selección de contenido se haga ni tampoco cómo se hace [Hamilton et al. 2014].

Bajo un sistema de competencia perfecta, en el que un gran número de vendedores ofrecen diversos servicios de VBD que compiten entre sí, y que cada comprador elige teniendo información completa sobre los méritos de cada uno de los servicios ofrecidos, cada comprador, tendría motivos para asumir que sus propios intereses como consumidor(a) son los que más influyen en el diseño de los servicios. En otras palabras, bajo competencia perfecta efectivamente "el cliente es el rey". Sin embargo, en un sistema de competencia imperfecta esto no siempre es así.

El tipo más "benigno" de manipulación es la experimentación corriente que realizan las plataformas para optimizar sus sistemas. Esta experimentación ocurre bajo la forma de *A/B testing* (pruebas A vs B) o *bucket testing* (pruebas por "cubetas" de usuarios), e involucra normalmente cambios de la interfaz que son probados con una pequeña cantidad de usuarios, como hacer un listado más largo o más corto, o cambiar el color, el tamaño, el orden o la posición de distintos elementos visuales. Si se observa que un cambio consistentemente incrementa la conversión o la satisfacción del usuario, observada directa o indirectamente, el cambio se hace permanente y se aplica a todos los usuarios. En un momento dado, puede haber decenas o cientos de condiciones experimentales que se están probando simultáneamente con distintos grupos de usuarios.

De alguna manera puede justificarse esta experimentación corriente cuando tiene el propósito de crear una mejor plataforma y ofrecer una experiencia más enriquecedora a los usuarios. Además, la plataforma, normalmente cuenta con el consentimiento de los usuarios para realizar estos experimentos. Por ejemplo, los términos de uso de Netflix indican que "analizamos continuamente varios aspectos de nuestro servicio" y permiten al usuario excluirse del *bucket testing* "visitando la página 'Cuenta' y cambiando la configuración de 'Participación en pruebas'" [Términos de uso de Netflix 2018].

En otras ocasiones, esta manipulación no es necesariamente en beneficio de los usuarios. El caso más conocido es el experimento de Facebook sobre la línea de tiempo de alrededor de 700.000 usuarios a los que se les enseñó contenido mayoritariamente positivo o mayoritariamente negativo durante varios días, para ver si eso influía en el tipo de contenido que esos usuarios publicaban. El resultado del experimento fue, efectivamente, comprobar que el estado de ánimo puede "contagiarse" de esta manera, y que los usuarios expuestos mayoritariamente a contenido negativo publicaban contenido más negativo y los usuarios expuestos a contenido positivo publicaban contenido más positivo *[*Kramer et al. 2014]. Cuando se dieron a conocer los resultados, la reacción de muchos usuarios de Facebook fue de indignación frente a lo que percibieron como una manipulación encubierta y realizada sin su consentimiento de su estado de ánimo.

Otro tipo de manipulación encubierta que no es en beneficio de los usuarios es la priorización de contenido producido por la propia plataforma sobre otros contenidos, una forma de integración vertical que finalmente resulta en desmedro de los consumidores. Hemos citado el caso de Google priorizando los resultados de sus propios servicios por encima de los de su competencia [Miller y Scott, 2014], y evidentemente en una plataforma de VBD se puede dar una situación análoga (Figura 11). Existen incentivos económicos para promocionar algunos contenidos sobre otros, incluso si éstos son menos relevantes para los usuarios y sin que los usuarios sepan que esta priorización está ocurriendo.

Figura 11: Cinco de las seis tragicomedias ofrecidas por este servicio en la primera pantalla son producciones propias

Finalmente, debemos recordar que, en principio, existe la capacidad tecnológica de realizar manipulación subrepticia con fines políticos o ideológicos, posiblemente de forma ilícita. El gran problema que tiene este tipo de manipulación en una plataforma de VBD es que es en gran medida invisible y difícil de demostrar: "La televisión puede ser monitoreada por cualquiera […] Si las noticias de la noche no cubren una protesta, la falta de cobertura es evidente […]. En cambio, no hay transparencia en el filtrado algorítmico: ¿cómo puede saber una persona si Facebook está

enseñando [noticias sobre una protesta] a todos los demás excepto a él o a ella, si simplemente no hay interés en el tema, o si se trata de un bucle de retroalimentación algorítmica, deprimiendo las actualizaciones a favor de un tema más amigable para el algoritmo …?" [Tufekci 2017].

3.2 Riesgos sobre la diversidad del contenido

La forma en que un medio de comunicación ofrece a una audiencia diversos tipos de contenido puede tener un impacto positivo o negativo sobre valores como la diversidad cultural y el pluralismo político/ideológico.

Respecto a la diversidad cultural en el contenido audiovisual, aspectos que tradicionalmente son de importancia para el regulador en Europa son la proporción de producción europea y contenidos de servicio público o interés general dentro de una oferta programática.

En el caso de las plataformas de VBD, diversidad no se refiere solamente al contenido que está disponible sino al contenido al que la audiencia es expuesta. Es decir, a la prominencia de estos contenidos, algo que es difícil de definir [Helberger et al. 2018]. La nueva revisión de la Directiva de Servicios de Comunicación Audiovisual (DSCAV) es un primer paso legislativo en la definición y aplicación de este principio [EU Parliament, 2018]. Tanto en sus considerandos 25 y 35, como en los artículos 7a y 13, se establece el principio de prominencia. En concreto, en el artículo 13 de la DSCAV, se indica que las plataformas de VBD "dispongan de un porcentaje de al menos el 30% de obras europeas en sus catálogos y garanticen la prominencia de dichas obras" y en su artículo 7a que debe haber una "adecuada prominencia" del contenido de interés general en estas plataformas.

Nos interesa saber si el contenido audiovisual muestra una imagen de la sociedad que de presencia y voz a distintos géneros y grupos franjas de edad, a gente de distinto origen, proveniente de diversas culturas, con distintas opciones religiosas o ideológicas y con capacidades variadas. Interesa también saber si personas distintas son representadas con roles sociales equivalentes o mediante estereotipos [Consell de l'Audiovisual de Catalunya 2013].

Respecto al pluralismo político/ideológico, algo particularmente preocupante puede ser la forma en que la selección de contenidos contribuye a crear "cámaras de resonancia" (*echo chambers*), llevando a un incremento de la polarización y el desconocimiento de otros puntos de vista distintos del propio producto de una "burbuja de filtrado" (*filter bubble*) creada por un sistema de recomendación [Parisier 2011].

3.2.1 Sobre-personalización

En la práctica, la personalización de contenidos ocurre en un continuo entre la personalización absoluta, en que las recomendaciones de cada persona son completamente independientes de las recomendaciones que reciben los demás, y la no-personalización, en la cual todas las personas reciben exactamente las mismas recomendaciones. En este sentido, una plataforma de VBD busca algún punto intermedio entre "la tiranía de la mayoría y las cámaras de resonancia" [Massa y Avesani 2007]. Técnicamente, se busca un equilibrio entre lo precisa que sea una recomendación por una parte, y lo diverso y novedoso que resulte el contenido recomendado por la otra [Landin et al. 2018].

La "tiranía de la mayoría" en la televisión abierta, donde sólo los contenidos que tienen una audiencia potencialmente grande son transmitidos, puede contraponerse con la posibilidad en una plataforma de VBD de acceder a contenido de nicho que difícilmente podría encontrar una audiencia en otras circunstancias. En teoría, estas plataformas podrían contribuir a la diversidad facilitando el descubrimiento de trabajos audiovisuales de bajo presupuesto o que no tienen una empresa de distribución o un presupuesto para publicidad: "gracias a los sistemas de recomendación, algunas películas podrían encontrar una audiencia incluso si no son parte de la programación de los canales de distribución convencionales" [European Audiovisual Observatory, 2018].

La promesa implícita del VBD es que los contenidos serán más relevantes para un usuario si los contenidos son escogidos de acuerdo con los intereses de esa persona. Por ejemplo, las condiciones generales de Movistar Plus indican que al recomendar contenidos a un usuario intenta "no volver a recomendarle contenidos que ya haya visualizado o que se separen radicalmente de sus hábitos de consumo" [Condiciones Generales de Movistar Plus, 2018]. Ésta promesa no es nueva, Nicholas Negroponte de hecho acuñó la expresión "*The Daily Me*" en los 1990s para referirse a la posibilidad de leer un periódico que sólo contendría noticias de interés para uno, con la consecuencia de que eventos potencialmente muy importantes podrían escapar completamente la atención de un usuario que no hubiese considerado la categoría de noticias correspondiente como relevante [Harper 1997].

El concepto de sobre-personalización (*over-personalization*) hace referencia a reducciones de la diversidad debido a un perfilado algorítmico demasiado estrecho, que introduce al usuario en una "burbuja" interpretando el interés que ha mostrado por un tipo de contenido como algo que **excluye** otros contenidos. En una situación así, es fácil caer en un bucle de retroalimentación en que, con el tiempo, los contenidos menos recomendados son también menos vistos reduciendo más aún, la diversidad de contenidos a los que un usuario está expuesto.

Estos bucles de retroalimentación a nivel individual pueden trasladarse a un fenómeno colectivo. Un ejemplo de ello podría ser el contenido documental o de cine independiente. "Dejando de lado el hecho de que Netflix y Amazon están comprando menos películas en general (lo cual también es un problema), y menos películas documentales y de cine arte cada año [...] cuando el algoritmo entierra esas películas, se convierte en un círculo vicioso" en el cual la plataforma puede decidir que este tipo de contenido "no funciona" y comprar aún menos al año siguiente [Newman 2017].

3.2.2 Personalización sesgada

Ningún sistema de recomendación, actual o previsible en el futuro cercano, es perfecto. Como hemos visto, se usan métodos estadísticos y probabilísticos para computar la probabilidad de que algo sea relevante para un usuario, sin ofrecer certidumbre. De alguna manera, esto es aceptable y aceptado por los usuarios mientras se trate de errores aislados y en alguna medida, fruto del azar. Cuando se trata de errores sistemáticos, hablamos de sesgo.

Un error sistemático puede ocurrir cuando el perfil que se ha creado de un usuario es erróneo, es decir, refleja intereses distintos de los que el usuario tiene. Esto puede ocurrir por varias razones. El caso más trivial es que, simplemente, más de una persona puede utilizar un mismo dispositivo sin necesariamente informar a una plataforma de VBD que se trata de un usuario distinto. En esta situación, la plataforma recibirá señales mezcladas provenientes de los intereses de varias personas

y puede erróneamente asociar un tipo de contenido al perfil de una persona para la cual ese tipo de contenido no es relevante.

Incluso si se trata de la persona correcta, la plataforma de VBD puede capturar variables generales acerca de un contexto (como el tipo de dispositivo usado, el lugar, o el momento en que se visualiza un contenido), pero no tiene suficiente información sobre cuáles son las razones que han llevado a un usuario a preferir un cierto contenido en un cierto contexto. "El usuario puede elegir ver *The Bridge*, *Top of the Lake* y *The Good Wife*. Puede haber un número de razones que hayan informado su decisión de mirar cada una, sin embargo, el sistema de recomendaciones personalizado [...] podría notar un patrón de protagonistas femeninas asertivas. Esto formará la base para nuevas recomendaciones, sometiendo al usuario a una identidad algorítmica" [Arnold 2016].

Otra forma de sesgo o error sistemático ocurre porque los sistemas de recomendación operan con datos de muchas personas al mismo tiempo y establecen correlaciones entre esos datos para realizar una recomendación a un individuo. En ese sentido, no están completamente exentos de la "tiranía de la mayoría" y pueden ser influenciados por correlaciones entre perfiles demográficos y de intereses compartidos por sexo o franja de edad, por ejemplo. Este efecto puede ser particularmente severo cuando un usuario utiliza por primera vez una plataforma (el "arranque en frío" o *cold start*) en los cuales esas correlaciones entre variables demográficas e intereses son prácticamente la única fuente de información de la que dispone un algoritmo.

Actualmente, existe la posibilidad para los usuarios en Europa de corregir estos perfiles erróneos. El artículo 21 del RGPD permite a un usuario no solamente conocer, corregir, y eliminar los datos que se hayan almacenado sobre él o ella, sino que también los perfiles de intereses que hayan sido automáticamente creados en base a esos datos. Naturalmente, una minoría relativamente pequeña de usuarios se toma la molestia de realizar estas correcciones.

4 Evaluando los efectos de las plataformas de VBD en la disponibilidad y acceso a contenidos

En esta sección introducimos una metodología para evaluar los efectos del uso de datos masivos en las plataformas de VBD sobre la disponibilidad y el acceso a contenidos audiovisuales. El punto de partida son los riesgos descritos en la sección anterior. Para cada una de las áreas de riesgo (autonomía y diversidad), presentamos una serie de factores que pueden ser observados. Luego, diseñamos una serie de indicadores para medir cada uno de los factores, y una metodología de observación para estos indicadores.

4.1 Factores a evaluar

La lista de factores que, en principio, pueden reflejar los valores que hemos descrito es potencialmente enorme. En esta sección, presentamos una propuesta de factores que son *relevantes* para cada área, y *distintos* en el sentido de que revelan aspectos diferentes del área.

4.1.1 Factores de autonomía

Los siguientes factores promueven la autonomía de los usuarios.

A1. Los usuarios saben que se les recomienda contenido automáticamente y de manera personalizada.
A2. Los usuarios conocen aspectos generales de cómo se genera una recomendación.
A3. Los usuarios conocen qué datos alimentan el proceso de recomendación, y pueden ver y corregir esos datos.
A4. Los usuarios conocen el perfil de intereses que se ha construido de ellos, y pueden ver y corregir dicho perfil.
A5. Los usuarios saben cuándo un contenido constituye un anuncio publicitario
A6. Los usuarios saben cuándo un contenido les ha sido recomendado en forma personalizada y cuando un listado ha sido ordenado en forma personalizada.
A7. Los usuarios pueden saber por qué les ha sido recomendado un contenido específico.
A8. Los usuarios pueden objetar una recomendación indicando que ese contenido no es relevante para ellos.
A9. Los usuarios disponen de mecanismos accesibles que les permiten encontrar un contenido específico.
A10. Los usuarios disponen de mecanismos no personalizados fácilmente accesibles para descubrir nuevos contenidos.

En esta lista de factores, contenidos, recomendaciones, y personalización se refieren a todo tipo de contenidos incluyendo el contenido publicitario y todos los anuncios que sean desplegados o enseñados a los usuarios.

4.1.2 Factores de diversidad

Los siguientes factores promueven la diversidad del contenido:

D1. Los contenidos ofrecidos a los usuarios incluyen contenidos que no son recomendados en forma personalizada.
D2. Los contenidos ofrecidos incluyen producción europea con una prominencia adecuada.
D3. Los contenidos ofrecidos incluyen contenido producido en distintos países.
D4. Los contenidos ofrecidos incluyen contenidos de servicio público o interés general con una prominencia adecuada.
D5. Los contenidos recomendados no excluyen completamente ninguna categoría, excepto en casos en que el usuario lo decida explícitamente.
D6. Los contenidos recomendados incluyen una mezcla de producción contemporánea y de períodos anteriores.
D7. Los contenidos recomendados incluyen elementos que no responden al estereotipo del perfil del usuario.

En esta lista de factores, "contenidos ofrecidos" se refiere a todo el contenido que es puesto a disposición de los usuarios, incluyendo el contenido recomendado, pero también contenido promovido por decisión editorial o porque es popular en un país o región. El contenido que no es recomendado de forma personalizada corresponde a contenido que es elegido mediante algún criterio que no es único para cada usuario, por ejemplo, por popularidad, recencia, puntuación por personas expertas u otros usuarios, o criterios editoriales. "*Inclusión*" y "*exclusión*" se refiere a que los contenidos estén fácilmente disponibles para los usuarios, no a que meramente formen parte de un catálogo.

4.2 Características deseables de indicadores y metodologías de observación

El siguiente paso es mapear cada factor a uno o varios indicadores que nos permitan evaluar hasta qué punto, un factor, está presente en una determinada plataforma de VBD. Estos indicadores pueden cubrir elementos cuantitativos y cualitativos y, para que sean útiles, idealmente deberían satisfacer al menos, los siguientes criterios:

- **Válido**: el indicador refleja correctamente un factor.
- **Accionable**: si el criterio asociado a un indicador no se cumple, los cambios necesarios para cumplirlo están claros
- **Distinto**: el indicador mide algo que no miden otros indicadores, o bien, lo mide de una manera distinta.
- **Observable externamente**: el indicador puede ser obtenido sin requerir la colaboración de las plataformas.

Estos indicadores deben ser analizados dentro de una cierta metodología de observación. Una metodología de observación debería cumplir al menos con los siguientes criterios:

- **Representativa**: en el caso de indicadores que pueden variar para distintos usuarios, la medida obtenida usando esta metodología es igual al promedio de lo que observarían los usuarios de la plataforma.

- **Eficiente en costo**: la metodología de observación no requiere una cantidad desmedida de recursos humanos o tecnológicos, lo que permite que pueda ejecutarse con una cierta periodicidad
- **Independiente de la plataforma**: la metodología de observación no requiere grandes modificaciones para ser aplicada a distintas plataformas, ni requiere grandes modificaciones si se realizan cambios menores en una plataforma
- **Resistente a la manipulación**: la plataforma no puede saber qué usuarios son observadores que miden indicadores, por lo que tanto no puede manipularlos mostrando una interfaz distinta, es decir, enseñando una interfaz diferente exclusivamente para estos observadores.

4.3 Indicadores propuestos

Proponemos los indicadores que aparecen a continuación para pedir los factores de autonomía y diversidad descritos anteriormente. Los indicadores de autonomía tienen tres valores posibles (presente, parcialmente presente, ausente). Los indicadores de diversidad tienen cuatro valores posibles (1, 2, 3, ó 4).

4.3.1 Indicadores de autonomía

Varios de los factores de autonomía reflejan estados internos de una persona que no son observables directamente, por ejemplo: si un individuo conoce o no, qué datos alimentan una recomendación. El énfasis en los indicadores propuestos es intentar evaluar aquello que cada plataforma de VBD provee y lo que podría cambiar, es decir, aquello que es *accionable* para la plataforma.

Para cada indicador de autonomía hay tres valores posibles: el indicador puede estar **presente, parcialmente presente, o ausente**. En caso de que un indicador esté parcialmente presente, cuando sea posible, se deben documentar los cambios mínimos que serían necesarios para determinar el indicador como completamente presente.

Tabla 2: Indicadores para cada factor de autonomía

	Factor	Indicadores
A1	Los usuarios saben que se les recomienda contenido automáticamente y de manera personalizada	● La plataforma comunica claramente a los usuarios que utiliza un sistema de recomendación automático la primera vez que los usuarios interactúan con la plataforma. ● Existen elementos de interfaz que recuerdan a los usuarios que hay un sistema de recomendación automático en operación.
A2	Los usuarios conocen aspectos generales de cómo se genera una recomendación	La plataforma informa a los usuarios acerca de cómo se genera una recomendación automática la primera vez que una recomendación personalizada es generada. La plataforma permite a los usuarios acceder fácilmente a una explicación que contiene aspectos generales de cómo se genera una recomendación.
A3	Los usuarios conocen qué datos alimentan el proceso de recomendación y pueden ver y corregir esos datos	La plataforma informa a los usuarios sobre qué datos serán usados para generar una recomendación personalizada al comenzar a capturar dichos datos La plataforma permite a los usuarios acceder fácilmente a un listado pormenorizado de cada tipo de dato utilizado para generar una recomendación La plataforma permite a los usuarios ver y corregir todos y cada uno de los datos que son utilizados para producir una recomendación
A4	Los usuarios conocen el perfil de intereses que se ha construido de ellos y pueden ver y corregir dicho perfil	La plataforma permite a los usuarios acceder al perfil de intereses que se ha creado de ellos La plataforma permite a los usuarios agregar, modificar, o eliminar algunos de estos intereses
A5	Los usuarios saben cuándo un contenido constituye un anuncio publicitario	La plataforma identifica de forma clara y evidente en la interfaz el contenido que es patrocinado o que constituye un anuncio publicitario
A6	Los usuarios saben cuándo un contenido les ha sido recomendado en forma personalizada y cuando un listado ha sido ordenado en forma personalizada	La plataforma identifica en forma clara y evidente en la interfaz el contenido que es recomendado automáticamente. La plataforma identifica de forma clara y evidente en la interfaz qué criterio se utiliza para ordenar una lista, incluyendo el caso de que el listado es ordenado por un criterio personalizado o de relevancia individual.
A7	Los usuarios pueden saber por qué les ha sido recomendado un contenido específico	La plataforma indica al usuario que hay información disponible acerca de por qué una recomendación ha sido realizada. La explicación provista por la plataforma es completa, comprensible.
A8	Los usuarios pueden objetar una recomendación indicando que ese contenido no es relevante para ellos	La plataforma provee al usuario de un mecanismo para objetar una recomendación. La plataforma provee al usuario de un mecanismo para evitar que recomendaciones similares sean enseñadas en el futuro
A9	Los usuarios disponen de mecanismos accesibles que les permiten encontrar un contenido específico	La plataforma provee de un mecanismo de búsqueda que permite encontrar un contenido específico.
A10	Los usuarios disponen de mecanismos no personalizados fácilmente accesibles para descubrir nuevos contenidos	La plataforma provee de menús de navegación y listados que despliegan contenido de manera no personalizada si el usuario lo solicita, sea por orden alfabético, por popularidad, o por criterios editoriales explícitos.

Notamos que la observación de estos indicadores debe **incluir contenido patrocinado y publicitario** respecto a los factores de autonomía A1-A8. Por ejemplo, el usuario debe saber que el contenido patrocinado y publicitario es escogido mediante un sistema automático, conocer su perfil para publicidad, modificarlo, saber por qué se le ha enseñado un cierto anuncio, indicar que tal anuncio no le resulta relevante, y así sucesivamente siguiendo los indicadores de A1-A8. Se excluyen aquí los factores de autonomía A9 y A10 que corresponden a buscar y descubrir contenido publicitario, funcionalidad que no resulta particularmente necesaria para los usuarios.

4.3.2 Indicadores de diversidad

Los indicadores de diversidad descritos a continuación se refieren principalmente al contenido que es ofrecido a los usuarios en dos vistas: la vista de inicio, que es la que el usuario ve primero cuando abre o inicia una plataforma y las vistas por categoría (por ejemplo, "Películas de Acción").

En estos indicadores, es importante el concepto de **prominencia** de un contenido que es ofrecido a un usuario, que como hemos mencionado está establecido en la nueva revisión de la DSCAV. La prominencia depende fundamentalmente de dónde y de qué forma es enseñado un contenido a un usuario.

Proponemos la siguiente **escala de prominencia**, que incluye tres niveles de prominencia desde el P1 al P3, siendo P1 el nivel más alto, además de NP (no prominente). Partimos del concepto de pantalla definido anteriormente.

- El nivel **P1 es el más alto**, y corresponde a contenido que está disponible en la primera pantalla de alguna forma, o en la segunda pantalla de forma destacada (mediante un color, borde o marca especial).
- El nivel **P2** está disponible en la segunda pantalla de alguna forma, o en la tercera pantalla de forma destacada.
- El nivel **P3** está disponible en la tercera pantalla.
- El nivel **NP** indica el contenido que no está disponible en ninguna de las tres primeras pantallas.

La prominencia depende del dispositivo usado, además del tamaño de pantalla en el caso de dispositivos disponibles en varios tamaños (como pantallas de ordenador o tabletas). En tal caso, la prominencia en una plataforma puede medirse en términos de pantallas físicas en relación a un dispositivo/pantalla o bien, de forma normalizada para una plataforma usando un concepto de **pantalla lógica**. Una pantalla lógica se define en términos de la cantidad de contenidos que caben en una pantalla física de una plataforma en un dispositivo de referencia (por ejemplo, un móvil o televisor inteligente que tenga una alta participación de mercado), o a un promedio de dispositivos de referencia. Por ejemplo, si el dispositivo de referencia es un ordenador portátil con una cierta resolución en la que caben 20 ítems por pantalla, los primeros 20 ítems son considerados parte de la primera pantalla lógica en cualquier otro dispositivo.

Para calcular cada indicador, primero se debe realizar un proceso de anotación del contenido desplegado en las primeras pantallas. Para cada indicador de diversidad hay cuatro valores posibles. El valor del indicador es el nivel (entre P1, P2, P3, recordando que P1 es más alto que P2 y P2 más

alto que P3) más alto necesario para encontrar el contenido descrito, o no prominente (NP) si el contenido no se encuentra en ninguno de los tres niveles de prominencia. Cada indicador se calcula para la página de inicio y para una página de vista por categorías específicas, o calculando el promedio de varias páginas de vistas por categoría.

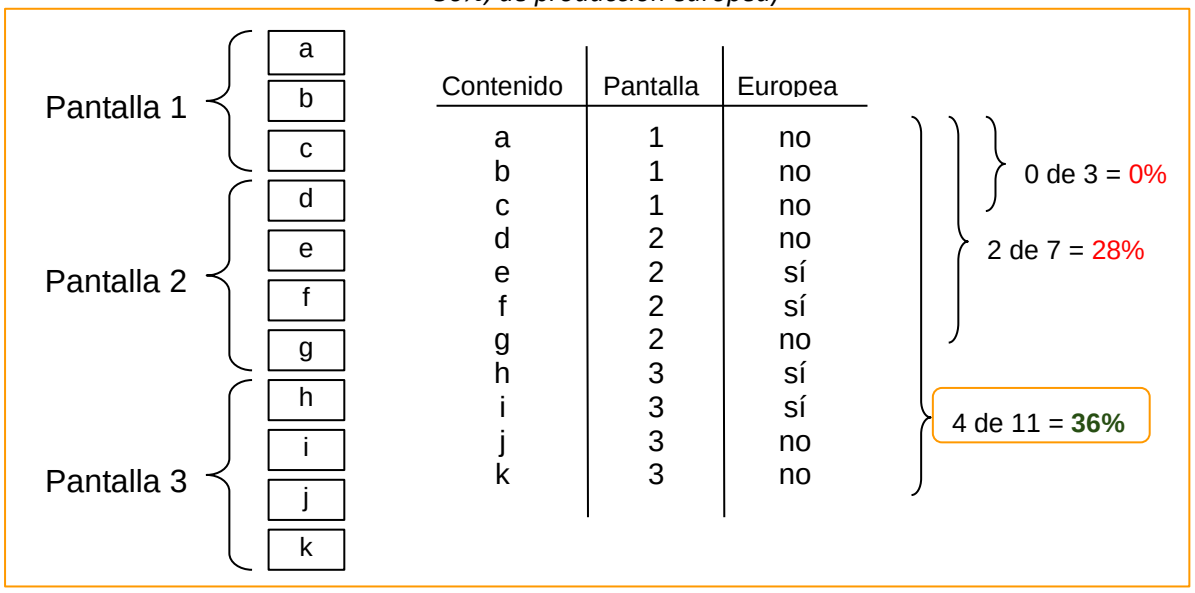
Tabla 3: Indicadores para cada factor de diversidad

	Factor	**Variable para anotación**	**Indicador**	**Mínimo propuesto**
D1	Los contenidos ofrecidos a los usuarios incluyen contenidos que no son recomendados de forma personalizada.	Variable binaria indicando si el contenido es recomendado o no.	Nivel más alto donde se encuentre al menos *X* contenidos no recomendados	1
D2	Los contenidos ofrecidos incluyen producción europea con una prominencia adecuada.	Variable binaria indicando si el contenido es de producción Europea o no	Nivel más alto donde se encuentre un *X%* de producción europea	30%
D3	Los contenidos ofrecidos incluyen contenido producido en distintos países.	Variable categórica indicando el país de producción	Nivel más alto donde se encuentren al menos *X* países distintos	2
D4	Los contenidos ofrecidos incluyen contenidos de servicio público o interés general con una prominencia adecuada.	Variable binaria indicando si el contenido es de servicio público o interés general	Nivel más alto donde se encuentre al menos *X* contenidos de servicio público o interés general (considerando categorías donde tengan cabida este tipo de contenidos)	1
D5	Los contenidos recomendados no excluyen completamente ninguna categoría, excepto en casos en que el usuario lo decida explícitamente.	Variable categórica indicando la macro-categoría o categoría general a la cual el contenido pertenece (largometrajes, ficción, formativos, documentales, deportes, ...)	Nivel más alto donde se encuentren *X* macro-categorías (Nota: se considera solamente la página de inicio)	3
D6	Los contenidos recomendados incluyen una mezcla de producción contemporánea y de períodos anteriores.	Variable ordinal indicando la década (relativa al año actual, no a decenios fijos) de producción del contenido	Nivel más alto donde se encuentren al menos *X* décadas (relativas al año actual, no a decenios fijos)	2
D7	Los contenidos recomendados incluyen elementos que no responden al estereotipo del perfil del usuario.	Variable ordinal indicando si el contenido corresponde al perfil de obvio/estereotípico de intereses del usuario	Nivel más alto donde se encuentre al menos *X* contenidos que no corresponda a este perfil	1

El único mínimo propuesto en la tabla anterior que está codificado explícitamente en la regulación es el 30% de producción europea. En general, los mínimos propuestos deberían ser ajustados y revisados en base a criterios desarrollados usando, en parte, la experiencia de aplicar esta metodología de observación a varias plataformas.

Ejemplo de cálculo de un indicador (Figura 12): supongamos que queremos calcular el indicador D2 para la página de inicio de una plataforma de VBD. Primero, se listan los contenidos disponibles en la primera, segunda, y tercera pantalla. Segundo, se anota cada contenido de acuerdo a si es producción europea o no. Tercero, se determina la proporción de producción europea en la primera pantalla, en la primera y la segunda pantalla, y en la primera, segunda y tercera pantalla. Cuarto, se observa cuál es el nivel más alto donde se encuentra un 30% o más de producción europea; el valor del indicador es ese nivel (P1, P2, P3), o NP si la condición no se cumple para ninguno de los tres niveles.

Figura 12: Ejemplo de cálculo de un indicador
(El valor del indicador D2 es P3 porque en las primeras tres pantallas hay un 36%, que es mayor que 30%, de producción europea)



4.4 Metodología de observación propuesta

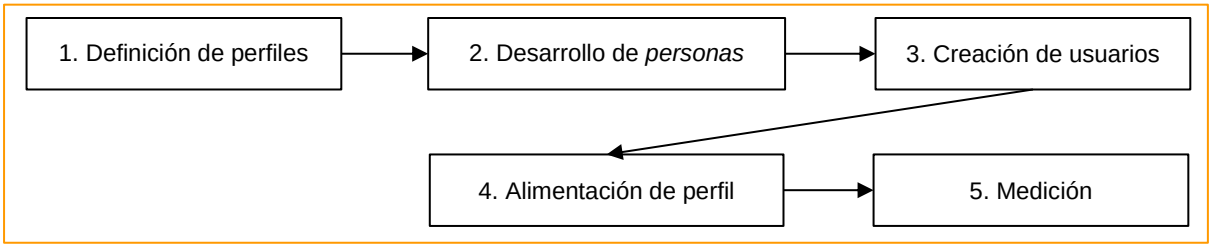
La metodología de observación que proponemos tiene **tres componentes que se complementan** entre sí:

- Captura de datos usando perfiles artificiales.
- Captura de datos usando perfiles naturales.
- Encuestas a usuarios de plataformas de VBD.

4.4.1 Captura de datos usando perfiles artificiales

Esta componente de la metodología de observación propuesta consiste en capturar datos usando perfiles artificiales o simulados de usuario (Figura 13):

Figura 13: Etapas de la captura de datos usando perfiles artificiales



Primero, se **definen una serie de perfiles representativos** del público que utiliza plataformas de VBD y vive dentro de un territorio. Estos perfiles pueden ser obtenidos mediante análisis del público del VBD, del público que consume audiovisual en general o bien, aproximados mediante perfiles sociodemográficos generales. En esta etapa deberían ser creados alrededor de 10 perfiles.

Segundo, se **desarrolla una *persona* para cada perfil** [Nielsen 2016] generando una narrativa ficticia que describe a una persona artificial. Esta narrativa incluye variables como edad, ocupación, educación, estado civil, lugar de residencia, trabajo, intereses respecto a distintos tipos de contenido, dispositivos digitales que posee, patrones de consumo audiovisual en términos de lugares, dispositivos, días y horarios en que se produce el consumo. Estas narrativas deberían representar a alguien que de alguna manera encarna a una persona prototípica dentro de cada perfil, y pueden incluir elementos exagerados o caricaturizados mientras sean creíbles.

Tercero, se **crea un usuario para cada perfil** en cada una de las plataformas a observar, pagando una suscripción mediante algún mecanismo de pago y utilizando una conexión a Internet que no sea trivialmente identificable como un posible observador/regulador de este mercado.

Cuarto, se **alimenta este perfil** comenzando por completar el perfil de usuario de cada plataforma de acuerdo con la *persona* desarrollada para ese perfil. Posteriormente, el perfil artificial debe usarse para visualizar/consumir contenido de acuerdo con los patrones de consumo e intereses determinados para la *persona*. Este consumo debe ser: (i) lo más representativo posible de la *persona* creada, (ii) realizado de la forma más natural posible, por ejemplo, desde los dispositivos que la *persona* tiene y, (iii) idealmente, en el mismo contexto (tiempo, lugar, dispositivo) en cada plataforma observada. El período durante el cual se alimenta un perfil debería extenderse durante

varios días (por ejemplo, 10-20 días) para permitir a cada plataforma capturar suficientes datos para la creación del perfil.

Quinto, se **realiza la medición** de los distintos indicadores desde un dispositivo asociado a cada perfil artificial. La observación para todos los perfiles debiese realizarse en el mismo día y dentro de la misma franja horaria. Se debe grabar una fotografía o captura de las pantallas usadas para calcular los indicadores.

La captura de datos usando perfiles artificiales es eficiente en costo e independiente de la plataforma, aunque puede ser menos representativa y menos resistente a la manipulación que la captura de datos usando perfiles naturales.

4.4.2 Captura de datos usando perfiles naturales

Este componente de la metodología de observación propuesta consiste en utilizar perfiles de usuario reales para capturar datos:

La primera etapa, **definición de perfiles representativos**, es idéntica al caso de los perfiles artificiales, aunque el número de perfiles puede ser mayor.

Segundo, para cada plataforma se **recluta un panel** de informadores que consiste en una cantidad de usuarios que correspondan con cada perfil (por ejemplo, 3-4 informadores por perfil, para que exista redundancia) y que hayan sido usuarios de esa plataforma de VBD durante un período mínimo como, por ejemplo, un año. El panel de usuarios puede estar constituido por voluntarios o bien, puede haber una compensación (no necesariamente monetaria). Debe haber una rotación de informadores, por ejemplo, que en promedio cada informador dure dos años. Para asegurar cierta continuidad, no se deberían cambiar todos los informadores a la vez, sino en forma escalonada.

Tercero, se establece un **mecanismo para proveer información** que resulte sencillo de ejecutar para cada informador. Este mecanismo puede ser, por ejemplo, hacer una serie de fotografías con el móvil a distintas pantallas de la interfaz y enviarlas a una dirección de correo electrónico.

Cuarto, se realiza un **seguimiento periódico** en el cual se solicita a los informadores que provean la información requerida con la periodicidad establecida; por ejemplo, una vez al mes.

Quinto, una persona experta **realiza la medición** de los distintos indicadores mediante la observación y anotación de la información provista por los informadores, por ejemplo, observando y anotando las fotografías enviadas por cada informador.

La captura de datos usando perfiles naturales es eficiente en costo, independiente de la plataforma, es más representativa de los usuarios y más resistente a la manipulación que la captura de datos usando perfiles artificiales. Existe complementariedad también: los perfiles artificiales permiten una observación más focalizada en aspectos específicos, mientras que los perfiles naturales permiten una observación más general de la experiencia de los usuarios de estas plataformas.

Si se realiza captura de datos usando perfiles artificiales y naturales, ambos procesos pueden compartir la etapa de medición, alimentándose con capturas de pantalla o fotografías realizadas en base a perfiles artificiales o naturales, y procesándolos de la misma manera. Complementariedad de metodologías.

4.4.3 Encuestas a usuarios de plataformas de VBD

Finalmente, las encuestas son una componente de la metodología de observación propuesta que complementa la captura de datos usando perfiles artificiales y/o naturales. Esto puede realizarse mediante una **encuesta específica sobre VBD** o bien, como **parte de una encuesta general sobre contenido audiovisual**. Cada factor medible mediante una encuesta, particularmente en el caso de los factores de autonomía, puede ser obtenido mediante una o varias preguntas.

Hay varios motivos por los cuales es necesario hacer esto. El primero y más importante es determinar sesgos o errores sistemáticos en la metodología de observación, contrastando los valores medidos para algunos factores entre la encuesta y la captura de datos artificial o natural. Esto permite, también, recolectar datos de una muestra de usuarios más grande mejorando así, la representatividad de la metodología.

Finalmente, el uso de una encuesta sobre VBD permite realizar preguntas abiertas que capturen nuevas amenazas a la autonomía de los usuarios o a la diversidad de los contenidos disponibles y que estos, permitan hacer evolucionar dicha metodología frente a cambios tecnológicos y sociales que lo requieran.

5 Plataformas para compartir vídeos y redes sociales

Las **plataformas para compartir vídeos** (PCV, o en inglés *VSP - Video Sharing Platforms*) son servicios que consisten en el alojamiento de gran cantidad de vídeos sobre los cuales la plataforma no tiene responsabilidad editorial, pero sí de ordenamiento, etiquetado, y posicionamiento mediante mecanismos principalmente algorítmicos, como hemos descrito. Las PCV tienen como principal propósito poner a disposición del público estos vídeos a través de redes electrónicas de comunicación. Ejemplos de estas plataformas son YouTube, Vimeo y DailyMotion.

Las plataformas para compartir vídeos no están reguladas en la DSCAV actualmente vigente [European Audiovisual Observatory, 2018], pero sí están integrados en la nueva revisión de la DSCAV [EU Parliament 2 Oct 2018]. La razón es que, al igual que las plataformas de VBD o de televisión lineal, también despliegan contenido audiovisual, a la vez que compiten básicamente por la misma audiencia y por los mismos ingresos publicitarios.

Las **plataformas de redes sociales** (RRSS) también caen bajo la regulación de la nueva Directiva, siempre y cuando se trate de plataformas "cuya oferta de programas y vídeos generados por usuarios constituya una funcionalidad esencial", dejando para futura regulación definir este carácter de esencialidad (DSCAV artículo 1a bis). Además, se incluyen los servicios que tengan una parte "disociable" que opere como una PCV o VBD. Se excluyen de la Directiva específicamente las plataformas que sirven solamente breves extractos de vídeos de periódicos y revistas, imágenes animadas y todo lo que ocurra en web privadas y comunidades no comerciales [EU Parliament 2 Oct 2018].

En esta sección extendemos a las PCV y RRSS el análisis que hemos hecho en relación con las plataformas de VBD (y al uso que hacen de datos masivos en la puesta a disposición de contenido audiovisual). Comenzamos describiendo cómo los riesgos que estas plataformas tienen para la autonomía del individuo y la diversidad del contenido son similares al caso de VBD, a la vez que hay nuevos riesgos para colectivos vulnerables que son específicos de PCV y RRSS (5.1). Posteriormente describimos una metodología de observación para PCV (5.2) y para RRSS (5.3), ambas derivadas de la metodología de observación para VBD.

5.1 Riesgos en las plataformas para compartir vídeos y en las de redes sociales

Las PCV y RRSS tienen características similares a las VBD además de características propias. Entre estas últimas las que impactan más directamente sobre la disponibilidad y el acceso a contenidos audiovisuales son la diferencia en la modalidad de acceso y la diferencia en el tipo de contenido.

Respecto a la **modalidad de acceso**, normalmente en PCV/RRSS no hay una barrera de pago (*paywall*) entre el contenido y los usuarios con lo cual, algunas de estas plataformas permiten acceso anónimo o seudónimo. La personalización en estos casos, o se realiza utilizando solamente variables contextuales generales como el país/ciudad desde donde el usuario visita o la hora del día o bien, se realiza utilizando un mecanismo de rastreo (*tracking*) de usuario mediante *cookies* que son asociadas a un navegador web. No siempre se provee acceso anónimo: en algunos casos la PCV/RRSS solo puede permitir acceso a través de una aplicación gratuita en la cual el usuario debe estar registrado, al menos, con una dirección de correo electrónico o un número de teléfono móvil.

Respecto al **tipo de contenido**, en PCV/RRSS encontramos en forma mayoritaria contenido generado por usuarios, cuya característica fundamental es que es "producido fuera de la rutina y práctica profesional". Es decir, de forma aficionada o no profesional, sin un mercado en mente y sin una expectativa de remuneración [Working Party on the Information Economy, 2006]. El contenido generado por usuarios abarca un espectro de calidad y temáticas vastamente más amplio que el contenido generado profesionalmente, y es puesto a disposición de los usuarios sin el control editorial característico de los medios tradicionales [European Audiovisual Observatory, 2018]. Esto permite que contenido de carácter no comercial, o que no está desarrollado para ser potencialmente monetizable a través de publicidad, pueda alcanzar una audiencia. El contenido generado por los usuarios plantea importantes oportunidades, como la posibilidad de permitir a una pluralidad de actores encontrarse y compartir contenido que promueva los intereses de comunidades minoritarias: la diáspora de distintas naciones; o bien minorías definidas por una orientación sexual, una etnia, un lenguaje, una situación de discapacidad, o en general, un contexto particular que los sitúe fuera del *mainstream* mediático.

A pesar de las diferencias en lo fundamental, los riesgos relacionados con la puesta a disposición de contenido audiovisual basándose en datos masivos, son los mismos que en el caso de las plataformas de VBD: riesgos para la autonomía individual y riesgos para la diversidad del contenido, a los que debemos agregar riesgos para colectivos vulnerables producto de la personalización de contenidos, tal como indicamos a continuación.

5.1.1 Riesgos para la autonomía individual y la diversidad del contenido

Hemos mencionado dos tipos de riesgos para la autonomía individual: la asimetría de información y la manipulación encubierta.

Respecto a la **asimetría de la información**, la situación de las plataformas de VBD se extiende naturalmente a las PCV/RRSS, porque es la plataforma la que conoce en detalle al usuario, mientras que el usuario no conoce en detalle a la plataforma. Por ejemplo, menos del 25% de los usuarios de Facebook saben que su línea de tiempo es resultado de un proceso de selección y filtrado, y una cantidad aún menor sabe cómo modificar este proceso [Hamilton et al. 2014]. Un elemento nuevo respecto a las plataformas de VBD es que la asimetría puede extenderse a los anunciantes en estas plataformas, quienes eventualmente pueden tener acceso a información personal sobre los usuarios. Esta práctica está cada vez más regulada.

Respecto a la **manipulación encubierta**, el riesgo es mayor que en las plataformas de VBD, ya que tanto las plataformas para compartir vídeos como las redes sociales son menos selectivas en los contenidos que ofrecen. Puede haber la presencia de terceros que intenten manipular a los usuarios, principalmente mediante nuevas estrategias publicitarias como por ejemplo el uso de sofisticadas herramientas de segmentación y focalización (*targeting*) de usuarios.

El caso reciente más importante que demostró este riesgo es el escándalo de *Cambridge Analytica*, empresa británica acusada, entre otras cosas, de haber utilizado Facebook para influir en el año 2016 en la elección presidencial en EE. UU. y el referéndum por el Brexit en el Reino Unido, violando la legislación en ambos países en materia de campañas electorales. Según investigaciones publicadas por el New York Times y The Guardian en marzo de 2018 [Cadwalladr 2018, Rosenberg 2018], *Cambridge Analytica* utilizó datos recopilados desde Facebook acerca de millones de personas. Con

ellos construyó modelos psicológicos de estos usuarios y utilizó estos modelos para sembrar en redes sociales propaganda electoral micro-focalizada en perfiles muy específicos.

La manipulación encubierta se vuelve más perjudicial cuando se combinan varios elementos, incluyendo el acceso subrepticio a datos personales, el conocimiento de expertos en comunicación política, y sofisticados algoritmos de inteligencia artificial usados para construir una maquinaria de "propaganda computacional" [Petit 2018a]. Esta maquinaria actúa tanto fabricando nuevas tendencias como silenciando voces disidentes, no solamente mediante la censura directa de ciertos contenidos, sino sembrando desinformación y ahogando esas voces en un mar de información basura creada por mecanismos automáticos [Tufekci 2017].

También hemos mencionado como amenazas a la diversidad producto del uso de datos masivos los riesgos que existen de **sobre-personalización** y **personalización sesgada**. En ambos casos, el hecho de que el contenido puede abarcar un espectro más amplio de temáticas y representar voces más diversas que en las plataformas de VBD plantea un riesgo aumentado de **polarización**. En RRSS se ha observado que, si bien hay una interacción entre distintas sensibilidades ideológicas y políticas, el contenido que se comparte tiende a no cruzar barreras ideológicas, lo que significa que un usuario está más expuesto a contenidos que refuerzan su posición que a contenidos que potencialmente la contradicen [Conover et al. 2011]. Esta polarización no es estática, sino que puede aumentar o disminuir con el tiempo; de hecho, se ha observado que períodos de mayor violencia política están correlacionados con períodos de mayor polarización en redes sociales [Weber et al. 2013].

La personalización de contenidos es la causa de las **burbujas de filtrado** ("*filter bubbles")*, en que mucho de lo que vemos en Internet confirma aquello en que ya creemos. Por ejemplo, una búsqueda de información sobre cambio climático puede dar resultados diametralmente opuestos dependiendo de si quien busca es una persona más interesada en la liberalización de la economía o en la protección del medio ambiente [Pariser 2011]. De la misma manera, la personalización puede convertir a una PCV en una herramienta de **radicalización**, en que una búsqueda por discursos de políticos moderados de algún signo político, por ejemplo, puede llevar rápidamente mediante una cadena de búsquedas recomendadas y vídeos recomendados a teorías de conspiración y discursos mucho más radicales, "como si el algoritmo hubiese concluido que la gente es atraída por contenido más extremo que el contenido que vieron inicialmente, o en general por contenido incendiario" [Tufekci 2018].

5.1.2 Riesgos de sobreexposición de colectivos vulnerables a contenido potencialmente dañino

Un riesgo específico de las PCV y RRSS respecto a las plataformas de VBD es que los mecanismos de personalización pueden resultar en una **sobreexposición de colectivos vulnerables a contenidos que sean dañinos para ellos** (incluyendo anuncios y contenido patrocinado).

Algunos ejemplos de colectivos de riesgo y tipos de contenido que puede resultar potencialmente perjudicial son los siguientes:

- **Personas con trastornos de alimentación**, anorexia, bulimia, o que hayan padecido de estos trastornos: contenidos que inciten o promuevan la anorexia (*pro-ana*, *thinspiration*) y otros trastornos alimentarios, o que promuevan una visión negativa del propio cuerpo.

- **Personas con riesgo de suicidio o autolesión**, que tengan algún factor de riesgo como una enfermedad mental, un trauma, un contexto psicosocial particular o un historial de intentos de suicidio o de autolesiones: contenidos que inciten o promuevan el suicidio o a la autolesión.
- **Personas con tendencia a la violencia**, o con un historial de comportamiento violento, sea esta violencia de género, homófoba, racista, xenófoba, por motivos religiosos o ideológicos y/o maltrato animal: contenidos que promuevan o inciten a actos violento, sea mediante información factualmente correcta o mediante campañas de desinformación o bien, que invite al usuario a unirse a grupos que usen la violencia.
- **Personas con drogodependencia** o rehabilitados de una drogodependencia, incluyendo el alcoholismo y el tabaquismo: contenidos que inciten al consumo de alcohol, tabaco, y otras drogas
- **Personas con ludopatía** o con tendencia o un historial de juego compulsivo: contenido que incite o promueva las apuestas.

Debemos hacer notar dos cosas. Primero, que evidentemente una lista de este tipo no puede ser exhaustiva porque las categorías de colectivos vulnerables a ciertos tipos de contenido evolucionan y cambian con el tiempo. Segundo, que con la información de la que dispone una plataforma es imposible determinar a ciencia cierta si una persona pertenece o no, a un grupo vulnerable: el diagnóstico de algunas de estas condiciones requiere entrevistas en profundidad con una persona experta. Sin embargo, aun cuando una plataforma no haga esta determinación correctamente en todos los casos o no la haga de forma explícita, la **consecuencia** de la personalización basada en datos puede ser la sobreexposición a un contenido perjudicial.

Para evaluar este riesgo, consideramos dos tipos de contenido respecto a cada comunidad vulnerable: contenido protector y contenido incitador. El **contenido protector** promueve una perspectiva o una serie de acciones que apuntan a mantener a una persona en riesgo alejada de aquello que le resulta perjudicial. Por ejemplo, en el caso del alcoholismo o la drogodependencia, un vídeo con consejos para evitar una recaída constituye contenido protector; en el caso de tendencias suicidas, información para contactar con profesionales en prevención del suicidio puede constituir contenido protector.

En cambio, el **contenido incitador** promueve una perspectiva o una serie de acciones que incitan a alguien en riesgo a exponerse a algo que le ha resultado perjudicial. Por ejemplo, en el caso de una persona en rehabilitación por alcoholismo, contenido que promueva el consumo de alcohol sería incitador. Notamos que ambos tipos de contenido, protector o incitador, puede aparecer tanto en el contenido principal de la plataforma como en el contenido publicitario.

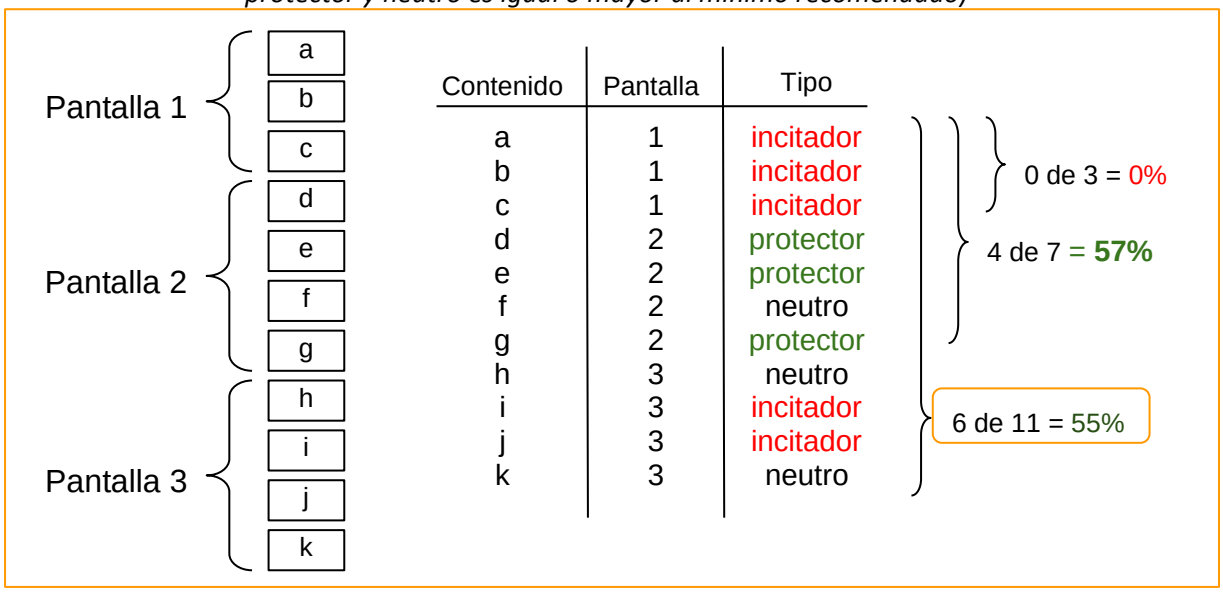
En base a esta distinción, proponemos dos factores con sus correspondientes indicadores, que también se computan en base a los niveles de prominencia (P1, P2, P3, NP):

Tabla 4: Indicadores para cada factor de riesgo para un colectivo vulnerable

	Factor	**Variable para anotación**	**Indicador**	**Mínimo propuesto**
C1	Los contenidos ofrecidos a usuarios con un perfil de riesgo incluyen contenido protector	Variable binaria indicando si el contenido es protector para un cierto riesgo o no.	Nivel más alto donde se encuentre al menos *X* contenidos protectores	1
C2	Los contenidos ofrecidos a usuarios con un perfil de riesgo incluyen menos contenido incitador que contenido neutro o protector	Variable categórica indicando si el contenido es incitador, protector, o neutro para un cierto riesgo o no	Nivel más alto donde el contenido protector y neutro sea más de X% del total (el total considera incitador, protector, y neutro)	50%

La Figura 14 muestra un ejemplo del cálculo de uno de estos indicadores.

Figura 14: Ejemplo de cálculo de un indicador de riesgo para colectivo vulnerable
(El valor del indicador C2, en este caso, es P2 porque al llegar a la segunda pantalla, el contenido protector y neutro es igual o mayor al mínimo recomendado)



5.2 Metodología de observación en plataformas para compartir vídeos

Proponemos para las PCV la misma metodología de observación de tres componentes que hemos propuesto para VBD, con las siguientes modificaciones:

1. considerar múltiples páginas de inicio,
2. crear perfiles artificiales de *personas* de colectivos vulnerables, y
3. medir los indicadores de riesgo para colectivos vulnerables.

La observación debe considerar que, en vez de una página de inicio que es lo normal en una plataforma de VBD, pueden existir **múltiples páginas de inicio** en una PCV. En caso de que existan múltiples páginas de inicio, la metodología de observación debiese incluir al menos:

(1) la página de inicio para un usuario que ha ingresado (*login*) a la plataforma, que es lo que más se aproxima a la página de inicio de una plataforma de VBD

(2) la página de suscripciones o la vista personalizada de inicio para este usuario si existe y es distinta de la página de inicio normal y

(3) la página de inicio para un usuario que no ha ingresado a la plataforma y que tiene bloqueada la focalización (*targeting*) precisa, por ejemplo, desactivando las *cookies* de un navegador. En este último caso, estamos observando la personalización que ocurre en base a variables generales de contexto como país, dispositivo, día de la semana u hora.

La observación debe incluir **perfiles artificiales de *personas* de colectivos vulnerables**, incluyendo las categorías descritas anteriormente, y **medir los indicadores de riesgo para colectivos vulnerables**. Los perfiles tendrán patrones de consumo correspondientes a *personas* que tengan los factores de riesgo descritos, por ejemplo, un perfil artificial que busca obsesivamente vídeos sobre métodos de suicidio. El uso de perfiles naturales de colectivos vulnerables plantea problemas éticos como la protección de las mismas personas en una situación de vulnerabilidad y, por lo tanto, se desaconseja como mecanismo de observación en este caso.

5.3 Metodología de observación en redes sociales

Proponemos para plataformas de RRSS la misma metodología de observación que para las PCV, con la diferencia de que debemos tener en cuenta la posible parte "disociable" de las mismas que es específica del contenido audiovisual [EU Parliament 2 Oct 2018].

Esto significa que la parte que se observará de las plataformas será el conjunto de páginas que presentan exclusivamente contenido audiovisual, solo en caso de que las plataformas lo tengan. Por ejemplo, Facebook actualmente provee de un servicio llamado "Facebook Watch" que opera básicamente como un híbrido entre una plataforma de VBD, con derechos sobre series y otros contenidos producidos comercialmente y una PCV basada en contenido provisto por los usuarios. A esta parte de la plataforma de red social se le puede aplicar la metodología de observación para PCV descrita anteriormente.

6 Recomendaciones

A continuación se presentan cuatro recomendaciones producto de este estudio:

1. **Monitorizar** los efectos del uso de datos masivos en la disponibilidad y acceso a contenidos audiovisuales
2. **Educar** respecto a derechos en materia de autonomía y diversidad y posicionarse a favor de la transparencia
3. **Identificar** los mecanismos para protección de menores usados en servicios audiovisuales
4. **Colaborar** con otras entidades respecto a contenido audiovisual ilegal

Las primeras dos se refieren al uso de algoritmos, que es el tema central de esta investigación. Las últimas dos se refieren a otros aspectos de las plataformas analizadas.

6.1 Monitorizar los efectos del uso de datos masivos en la disponibilidad y acceso a contenidos audiovisuales

Una serie de nuevos servicios están teniendo un rol cada vez más importante en el acceso a contenidos audiovisuales. La proporción de contenidos a los que una persona tiene acceso mediado por algoritmos basado en datos masivos sólo puede esperarse que aumente con el tiempo. Es el momento de comenzar a monitorizar los efectos que están teniendo estos algoritmos, considerando particularmente los riesgos que hemos descrito.

La primera recomendación es poner en práctica en forma escalonada una metodología de monitorización, comenzando con una prueba inicial que incluya solamente dos o tres plataformas de Vídeo Bajo Demanda que tengan el mayor número de suscriptores, con captura de datos mediante tres o cuatro perfiles artificiales, y usando como dispositivo de referencia un único Smart TV que tenga alta participación de mercado. Esta metodología puede luego expandirse a más plataformas de VBD, más perfiles, y más dispositivos de referencias. Luego puede incorporarse la monitorización de Plataformas para Compartir Vídeos y también a los elementos específicos de las plataformas de Redes Sociales para el consumo de contenido audiovisual; al tiempo que se vayan agregando más fuentes de información como perfiles naturales y encuestas.

La forma en que se monitoriza por primera vez de forma sistemática un nuevo servicio debe estar caracterizada por la flexibilidad para poder ir ajustando la metodología mediante la experiencia práctica. La estabilidad requerida para tener validez longitudinal es un objetivo importante, pero probablemente no puede ser alcanzado sin antes realizar ajustes razonables a la metodología, incluyendo modificaciones a los factores e indicadores descritos y a los mínimos propuestos que provengan de su puesta en práctica.

De cara al futuro, la estrategia de monitorización que hemos descrito es una forma preliminar de **auditoría algorítmica**. Se trata de una inspección independiente que busca comprobar si el funcionamiento de unos algoritmos (en este caso de búsqueda, recomendación y personalización de contenidos audiovisuales) se ajustan a unos parámetros determinados. Si se considera necesario en algún momento profundizar esta auditoría, muy probablemente se requieran cambios legales que posibiliten la colaboración de las plataformas además de desarrollos tecnológicos que incrementen la escalabilidad y, por lo tanto, el detalle y la frecuencia de las observaciones.

6.2 Educar respecto a derechos en materia de autonomía y diversidad y posicionarse a favor de la transparencia

La regulación en Europa en materia de privacidad y protección de datos mediante el RGPD y en materia de difusión audiovisual mediante la nueva DSCAV, traen una serie de **nuevos derechos** que es muy probable que estos sean desconocidos para los usuarios. Por ejemplo, el derecho que tienen las personas a acceder, modificar, o borrar datos de comportamiento que hayan sido recolectados automáticamente, el derecho a examinar y corregir los perfiles que la plataforma puede haber construido en base a estos datos o incluso, la portabilidad de datos entre plataformas de VBD.

Simultáneamente al avance de la regulación, hay un llamado cada vez más imperativo a que exista transparencia respecto a los servicios que median la información que recibimos. Por ejemplo, Angela Merkel ha llamado a las grandes plataformas de Internet a que revelen a los usuarios cómo y sobre qué bases se decide qué información las plataformas les enseñan [Connolly 2016]. La transparencia se hará cada vez más importante si se continúa difuminando la distinción entre la "realidad" producida de forma algorítmica y el "mundo" del cual supuestamente emana [Rouvroy 2013].

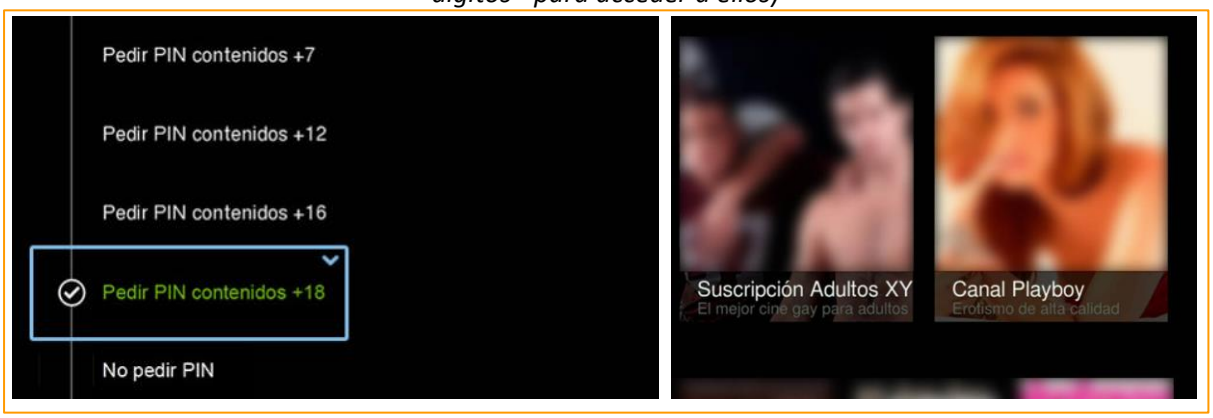
La segunda recomendación es promover la transparencia y educar a los usuarios respecto a sus derechos. Ambas cosas están íntimamente relacionadas, puesto que, por ejemplo, la protección de la autonomía individual requiere saber cómo se produce una recomendación, mientras que la diversidad de contenidos solamente puede ser evaluada en su totalidad si se conoce completamente el conjunto del cual una oferta de contenidos ha sido extraída.

Lograr transparencia algorítmica requiere contestar numerosas preguntas, incluyendo qué constituye una explicación, qué elementos pueden ser transparentes y cómo los requerimientos de transparencia pueden afectar el comportamiento de las plataformas [Fields et al. 2018]. La transparencia es básica para que puedan ejercerse los otros derechos respecto a un algoritmo [IEEE EAD 2018, ACMUS 2018] y la no-transparencia abre una oportunidad real a que se desarrollen y consoliden prácticas de manipulación encubierta.

6.3 Identificar los mecanismos para protección de menores usados en servicios audiovisuales

Actualmente la determinación de si un usuario es o no menor de edad, la realiza el titular de la cuenta de una plataforma de VBD o, simplemente, lo declara el mismo usuario de las PV y RRSS: no se usa un algoritmo ni técnicas de datos masivos para hacer esta inferencia (Figura 15).

Figura 15: Mecanismo para protección de menores
(Esta plataforma que ofrece contenido para adultos, pero requiere de un "Pin Parental" - código de 4 dígitos - para acceder a ellos)



Existe por lo tanto una primera responsabilidad de los padres o tutores legales de un menor que son los que le facilitan los medios materiales de acceso a estos contenidos. También las plataformas deben implementar medidas de protección que reduzcan el riesgo de que menores de edad tengan acceso a contenidos no aptos para ellos.

La tercera recomendación es identificar las medidas que actualmente toman las principales plataformas de VBD, PCV, y RRSS para reducir el riesgo de que menores de edad accedan a contenidos audiovisuales no aptos para ellos. Esta observación, que no requiere la colaboración de estas plataformas, debería considerar tanto la forma en que la plataforma determina que un usuario es menor de edad, como el contenido que le enseñan a este usuario una vez que la minoría de edad ha sido establecida. En base a lo observado, se deberían proponer una serie de buenas prácticas.

6.4 Colaborar con otras entidades respecto a contenido audiovisual ilegal

La monitorización del contenido audiovisual ofrecido a través de plataformas de VBD, de PCV y de RRSS con elementos específicos audiovisuales, además del monitoreo de contenido perjudicial para colectivos vulnerables y la identificación de medidas de protección de menores pueden generar subproductos positivos. Entre ellos se encuentra la posibilidad de realizar hallazgos de contenido audiovisual ilegal o contenido que ponga en riesgo a menores de edad y colectivos vulnerables.

Por otra parte, la metodología de observación que se ha descrito considera como punto de partida un usuario de la plataforma el cual se han capturado datos de comportamiento con el propósito de personalizar una oferta de contenidos. El énfasis, por lo tanto, no está en si el catálogo de la plataforma **contiene** o no un tipo de contenido sino en la forma en que este contenido es **puesto a disposición** de un usuario.

Monitorizar si la plataforma incluye o no contenido ilegal, incluyendo los anuncios, requiere inspeccionar el catálogo de contenido y el de anuncios, no la puesta a disposición de estos contenidos y anuncios. Esto necesita una metodología completamente distinta que no está basada en perfiles artificiales o naturales, sino en procedimientos automáticos de búsqueda y navegación para descubrir si un tipo de contenido o anuncio está o no presente. Además, las consecuencias de descubrir este contenido no son las mismas que las de descubrir una disminución de autonomía individual, de diversidad de contenidos o de sobreexposición de colectivos vulnerables a contenidos dañinos. Si se descubre contenido ilícito se debe tomar una acción que dependerá de la naturaleza de dicho contenido. No es lo mismo descubrir que con ciertas palabras clave aparecen anuncios publicitarios de tabaco que descubrir pornografía infantil o contenido que promueva el tráfico de armas o personas, que entran dentro del ámbito policial y judicial.

La cuarta recomendación es desarrollar colaboraciones con otras entidades respecto a contenido ilegal que establezcan sistemas de monitorización conjunta de distintas tipologías de contenido ilícito, utilizando los medios tecnológicos apropiados para realizar esta monitorización, incluyendo exploración y búsqueda automatizada que ayude a descubrir este contenido. Esta monitorización debería cubrir un espectro amplio de plataformas, incluyendo las que no son tan populares, debería realizarse con una frecuencia relativamente alta y debería contemplar prioridades y acciones específicas a realizar respecto a cada tipo de contenido ilícito que pueda ser encontrado.

7 Agradecimientos

El autor agradece la colaboración de Yasmina Soto de la Universitat de Barcelona y de los expertos del Consell de l'Audiovisual de Catalunya en la elaboración de este informe, indicando que todos los posibles errores y omisiones de este son de mi exclusiva responsabilidad.

8 Bibliografía

ACM US. Statement on Algorithmic Transparency and Accountability, 2017. Disponible en https://www.acm.org/articles/bulletins/2017/january/usacm-statement-algorithmic-accountability

ANDREOU, A; VENKATADRI, G; GOGA, O; GUMMADI, K. P; LOISEAU, P; MISLOVE, A. "Investigating ad transparency mechanisms in social media: A case study of Facebook's explanations." A The Network and Distributed System Security Symposium (NDSS), 2018.

ARNOLD, S. "Netflix and the Myth of Choice/Participation/Autonomy". A McDonald, K; Smith-Rowsey, D (Ed.) The Netflix Effect: Technology and Entertainment in the 21st Century. Publisher: Bloomsbury, 2016.

BAEZA-YATES, R; RIBEIRO-NETO, B. Modern information retrieval. Addison-Wesley Professional; 2nd edition, 2011.

BELL, R. M; KOREN, Y. "Lessons from the Netflix prize challenge."A Acm Sigkdd Explorations Newsletter, 9(2), 75-79, 2007.

BUNDESMINISTERIUM FÜR BILDUNG UND FORSCHUNG. Zukunftsprojekt Industrie 4.0 - BMBF. BMBF, January 21, 2016. Disponible https://www.bmbf.de/de/zukunftsprojekt-industrie-4-0-848.html

BROADBAND TV NEWS. Europeans watch two more hours a month On-Demand video, August 2018. Disponible en https://www.broadbandtvnews.com/2018/08/13/europeans-watch-two-more-hours-a-month-on-demand-video/

CADWALLADR, C; GRAHAM-HARRISSON, E. Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach. The Guardian, 17 març 2018. Disponible en https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election?CMP=Share_AndroidApp_Tweet

CHHABRA, S. "Netflix says 80 percent of watched content is based on algorithmic recommendations". Mobilesyrup, Aug 22, 2017. Disponible en https://mobilesyrup.com/2017/08/22/80-percent-netflix-shows-discovered-recommendation/

CHANDRASHEKAR, A; AMAT, F; BASILICO, J; JEBARA, T. Artwork Personalization at Netflix. Artwork Personalization at Netflix. Netflix Tech Blog, December 2017.

CHANNEL 4. Data, Democracy, and Dirty Tricks. News, 2017. Disponible: https://www.channel4.com/news/data-democracy-and-dirty-tricks-cambridge-analytica-uncovered-investigation-expose

CONNOLLY, K. "Angela Merkel: internet search engines are 'distorting perception'". The Guardian, October 27, 2016. Disponible en https://www.theguardian.com/world/2016/oct/27/angela-merkel-internet-search-engines-are-distorting-our-perception

CONOVER, M.; RATKIEWICZ, J; FRANCISCO, M. R; GONÇALVES, B; MENCZER, F; FLAMMINI, A. Political polarization on Twitter. ICWSM, 133, 89-96, 2011. Disponible en https://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/download/2847/3275

CONSEIL SUPÉRIEUR DE L'AUDIOVISUEL. Les mutations de la mise à disposition de contenus audiovisuels à l'ère du numérique: conséquences et enjeux - Le rôle des données et des algorithmes dans l'accès aux contenus. France, 2017. Disponible en http://www.csa.fr/Etudes-et-publications/Les-etudes-thematiques-et-les-etudes-d-impact/Les-publications-du-CSA-Lab/Les-mutations-de-la-mise-a-disposition-de-contenus-audiovisuels-a-l-ere-du-numerique-consequences-et-enjeux-Le-role-des-donnees-et-des-algorithmes-dans-l-acces-aux-contenus

CONSELL DE L'AUDIOVISUAL DE CATALUNYA. Informe sobre la diversitat i la igualtat a la televisió. Informe 58/2013. Barcelona: CAC, 2013.

CONSELL DE L'AUDIOVISUAL DE CATALUNYA. Fake news, algoritmes i bombolles informatives. Quaderns del CAC 44, Vol. XXI. Barcelona: CAC, 2018. Disponible en https://www.cac.cat/documentacio/fake-news-algoritmes-i-bombolles-informatives

CONSELL DE L'AUDIOVISUAL DE CATALUNYA. Les fake news a internet. El discurs de gènere. Informe 13/2018. Barcelona: CAC, 2018. Disponible en https://www.cac.cat/sites/default/files/fakenews.pdf

DANCE, G.J.X; LAFORGIA, M; CONFESSORE, N. As Facebook Raised a Privacy Wall, It Carved an Opening for Tech Giants. New York Times, Desembre 2018. Disponible en https://www.nytimes.com/2018/12/18/technology/facebook-privacy.html

EUROPEAN AUDIOVISUAL OBSERVATORY. The legal framework for video-sharing platforms. IRIS Plus, 2018. Disponible en https://rm.coe.int/the-legal-framework-for-video-sharing-platforms/16808b05ee

EUROPEAN BROADCASTING UNION. EBU reply to the European Commission Green Paper – Preparing for 2a Fully Converged Audiovisual World: Growth, Creation and Values, 2013. Disponible en https://www.ebu.ch/files/live/sites/ebu/files/Publications/Position papers/EBU-Position-EN reply Green Paper convergence 10.9.2013.pdf

EUROPEAN PARLIAMENT. Provision of audiovisual media services. European Parliament legislative resolution of 2 October 2018 on the proposal for a directive of the European Parliament and of the Council amending Directive 2010/13/EU P8_TA-PROV(2018)0364. (Particularment respecte a plataformes per a compartir videos i xarxes socials: considerants 1, 4, 5, 6, 45; Article 1(b), definicions). Disponible en http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+TA+P8-TA-2018-0364+0+DOC+XML+V0//EN

FIELDS, B; JONES, R; COWLISHAW, T. The Case for Public Service Recommender Algorithms. Invited talk at FATREC, 2018. Disponible en https://www.intotheminds.com/blog/en/fatrec-2018-keynote-on-public-service-recommendation-algorithms/

HAMILTON, K; KARAHALIOS, K; SANDVIG, C; ESLAMI, M. “A path to understanding the effects of algorithm awareness.” A CHI'14 Extended Abstracts on Human Factors in Computing Systems. pp.631-42. ACM, 2014.

HARPER, C. “The Daily Me”. American Journalism Review, abril 1997. Disponible en http://ajrarchive.org/Article.asp?id=268

HELBERGER, N; KARPPINEN, K; D’ACUNTO, L. “Exposure diversity as a design principle for recommender systems”. Information, Communication & Society, 21(2):191-207, 2018. Disponible en https://www.tandfonline.com/doi/full/10.1080/1369118X.2016.1271900

HILBERT, M; LÓPEZ, P. “The world's technological capacity to store, communicate, and compute information”. Science, 332(6025), pp. 60-65. Disponible en http://science.sciencemag.org/content/332/6025/60

IEEE EAD - Ethically Aligned Design v2, 2018 Disponible en https://ethicsinaction.ieee.org/

KRAMER, A. D; GUILLORY, J. E; HANCOCK, J. T. Experimental evidence of massive-scale emotional contagion through social networks. Proceedings of the National Academy of Sciences, 2014.

KOENE, A; PEREZ, E; CARTER, C. J; STATACHE, R; ADOLPHS, S; O’MALLEY, C; MCAULEY, D. “Ethics of personalized information filtering”. A International Conference on Internet Science, 123-32. Springer, Cham, 2015.

LANDIN, A; SUÁREZ-GARCÍA, E; VALCARCE, D. “When Diversity Met Accuracy: A Story of Recommender Systems”. Multidisciplinary Digital Publishing Institute Proceedings, 2(18), 1178, 2018.

LEW, M. S; SEBE, N; DJERABA, C; JAIN, R. “Content-based multimedia information retrieval: State of the art and challenges”. A ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), 2(1), pp. 1-19, 2016.

LI, L; CHU, W; LANGFORD, J; SCHAPIRE, R. E. “A contextual-bandit approach to personalized news article recommendation”. A Proceedings of the 19th international conference on World wide web (pp. 661-670). ACM, Abril 2010.

LOWENSOHN, J. "The science behind Netflix’s first major redesign in four years." The Verge, May 22, 2015 Disponible en: https://www.theverge.com/2015/5/22/8642359/the-science-behind-the-new-netflix-design

MACKENZIE, I; MEYER, C; NOBLE, S. How retailers can keep up with consumers. Octubre de 2013. Disponible en https://www.mckinsey.com/industries/retail/our-insights/how-retailers-can-keep-up-with-consumers

MASSA, P; AVESANI, P.“Trust metrics on controversial users: Balancing between tyranny of the majority and echo chambers”. International Journal on Semantic Web and Information Systems, 3(1), p. 39-64, 2007. Disponible en www.igi-global.com/article/trust-metrics-controversial-users/2830

MILLER, C.C; SCOTT, M. Google Settles Its European Antitrust case; Critics Remain, the New York Times, February 5th, 2014.

NEWMAN, B. Why Netflix and Amazon Algorithms Are Destroying the Movies, 2017. Disponible en https://www.indiewire.com/2017/07/netflix-amazon-algorithms-destroying-the-movies-1201853974/

NIELSEN, L. "Personas. Chapter 30". A The Encyclopedia of Human-Computer Interaction, 2nd Ed., 2016. Disponible en https://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed/personas

O'NEIL, C. Weapons of Math Destruction. Crown Random House, 2016.

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT. Working Party on the Information Economy - participative web: user-created content, 2006. Disponible en: http://www.oecd.org/internet/ieconomy/38393115.pdf

PARISER, E. The filter bubble: What the Internet is hiding from you. Penguin UK, 2011.

PASQUALE, F. The Black Box Society. Harvard University Press, 2015. Disponible enhttp://www.hup.harvard.edu/catalog.php?isbn=9780674368279

PETIT, M. "La política lingüística del segle XXI: Internet i la prominència dels continguts en català en els serveis audiovisuals a demanda". Revista de Llengua i Dret, Journal of Language and Law, (70), 1-16, 2018. Disponible en http://revistes.eapc.gencat.cat/index.php/rld/article/view/10.2436-rld.i70.2018.3186

PETIT, M. "Per una crítica de la raó algorítmica. Estat de la qüestió sobre la intel·ligència artificial, la seva influència en la política i la seva regulació". Quaderns del CAC 44, vol. XXI - juliol 2018(a). Disponible en https://www.cac.cat/sites/default/files/2018-08/Q44_CA_0.pdf

RADER, E; GRAY, R. "Understanding user beliefs about algorithmic curation in the Facebook news feed". A Proceedings of the 33rd annual ACM conference on human factors in computing systems. pp.173-82, 2015. Disponible en http://www.bierdoctor.com/papers/filterbubble_CHI2015_final.pdf

RADIO TELEVISIÓN ESPAÑOLA. Big Data, Conviviendo con el Algoritmo. Documents TV, Juliol 2017. Disponible en http://www.rtve.es/alacarta/videos/documentos-tv/documentos-tv-big-data-conviviendo-algoritmo/3893978/

REISMAN, D; SCHULTZ, J; CRAWFORD, K; WHITTAKER, M. Algorithmic Impact Assessments: A practical framework for public agency accountability. AI Now Institute, Abril 2018. Disponible en https://ainowinstitute.org/aiareport2018.pdf

ROSENBERG, M; CONFESSORE, N; CADWALLADR, C. How Trump Consultants Exploited the Facebook Data of Millions. The New York Times, 17 de març de 2018. Disponible en https://mobile.nytimes.com/2018/03/17/us/politics/cambridge-analytica-trump-campaign.html?referer=https://t.co/Hc9TwovbpO

ROUVROY, A. "The end (s) of critique: data behaviourism versus due process". A Privacy, Due Process and the Computational Turn: The Philosophy of Law Meets the Philosophy of Technology. pp. 157-182. Taylor & Francis, 2013.

SCHWAB, P.N. Special Talk: European Public Broadcasters Path towards Public Service Recommender Systems. FATREC, 2018. Disponible en https://www.intotheminds.com/blog/en/fatrec-2018-keynote-on-public-service-recommendation-algorithms/

SMITH, M; PATIL, D.J; MUÑOZ, C. Big Risks, Big Opportunities: the Intersection of Big Data and Civil Rights. US Office of the President, 2016. Disponible en https://obamawhitehouse.archives.gov/blog/2016/05/04/big-risks-big-opportunities-intersection-big-data-and-civil-rights

TUFEKCI, Z. Twitter and tear gas: The power and fragility of networked protest. Yale University Press, 2017.

TUFEKCI, Z. Youtube, the Great Radicalized. New York Times (Opinion), 2018. Disponible en: https://www.nytimes.com/2018/03/10/opinion/sunday/youtube-politics-radical.html

UNIÓN EUROPEA. Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE. Diario Oficial de la Unión Europea, L119/1 de abril de 2016 pp. 1-88. Disponible en https://eur-lex.europa.eu/legal-content/ES/ALL/?uri=CELEX:32016R0679

WEBER, I; GARIMELLA, V. R. K; BATAYNEH, A. "Secular vs. islamist polarization in egypt on twitter". A Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. pp. 290-297, 2013. Disponible en https://users.ics.aalto.fi/kiran/content/Secular-vs.-Islamist-Polarization-in-Egypt-on-Twitter.pdf

Versiones de los términos de uso utilizadas, revisadas en noviembre de 2018:

- HBO (Terminos de Uso y Condiciones): https://es.hboespana.com/terms-and-conditions
- HBO (Politica de Datos Personales) https://es.hboespana.com/data-policy
- Movistar Plus, Condiciones Generales: http://www.movistar.es/rpmm/estaticos/residencial/fijo/banda-ancha-movistarTV/condicionesgenerales-movistar-plus.pdf
- Netflix (Términos de Uso): https://help.netflix.com/legal/termsofuse
- Netflix (Personal Information Page): https://help.netflix.com/en/node/100624

Las capturas de pantalla en este informe han sido realizadas en los meses de noviembre y diciembre de 2018 por usuarios en Barcelona y/o Utrecht (Holanda), o son provistas por las mismas plataformas.