

MINERÍA DE DATOS EN SOCIAL MEDIA PARA EL BIEN PÚBLICO?

Tomado de “Post, Mine, Repeat, Social Media Data Mining Becomes Ordinary”, Helen Kennedy, Palgrave Macmillan, UK, 2016 (Capítulo 4).

Usos de la minería de datos en redes sociales

A través de las entrevistas realizadas en cinco organizaciones del sector público en 2012, descubrimos que se estaban utilizando las redes sociales y las herramientas de minería de datos enumeradas en la (Tabla 4.1) a continuación.

Table 4.1 Social media platforms and data mining tools used by public sector organisations in which we carried out interviews

<i>Social media</i>	<i>Data mining tools</i>
<ul style="list-style-type: none">• Twitter• Facebook• YouTube• Flickr• AudioBoo• Instagram• Get Glue• Pinterest• Foursquare• blogs	<ul style="list-style-type: none">• Google Analytics• Museum Analytics• Google Alerts• Rate This Page functions• Postling• Hootsuite• One Riot• Site Improve• Storify• Crowdbooster• iTunes download monitoring• internal Facebook/Twitter tools

Estas herramientas se utilizaron con diversos grados de regularidad. Todas las organizaciones tenían cuentas en varias plataformas y ninguna las mantenía todas regularmente. En algunas organizaciones, una plataforma se usaba más regularmente y otras más esporádicamente; en otros casos, una cuenta particular de Twitter, por ejemplo, estaría especialmente activa, mientras que otras cuentas se utilizaron con menos frecuencia. Los entrevistados sintieron que sus organizaciones carecían de recursos suficientes, lo que significaba que dedicaban recursos limitados a la participación y el análisis de las redes sociales.

Algunos lograron administrar y monitorear las redes sociales a pesar de los recursos limitados, debido a su estructura a pequeña escala o autónoma, mientras que otros querían hacer más con los recursos limitados que tenían. Algunas organizaciones transmitieron datos cuantitativos a los financieros o gerentes superiores, utilizando herramientas como Museums Analytics para producir resúmenes de la actividad de las redes sociales, como nuevos ‘Me gusta’ de la página, publicaciones y comentarios en Facebook, y nuevos seguidores, tweets o menciones en Twitter.

Los entrevistados no sabían qué sucedió con estos datos, si es que sucedió algo, y algunos de nuestros entrevistados cuestionaron la medida en que los datos cuantitativos podrían medir el compromiso "real", creyendo en cambio que se debería poner más énfasis en los datos cualitativos. Otros entrevistados consideraron que era

difícil persuadir a los altos directivos del valor y los beneficios del monitoreo de las redes sociales. Para demostrar la comprensión limitada de la cultura de las redes sociales que existía entre los colegas superiores, un entrevistado nos dijo en broma que su gerente pensaba que LinkedIn se llamaba LinkedUp.

Uno de los gobiernos locales supervisó las redes sociales tanto para el servicio al cliente como para la gestión de la reputación, y las acciones identificables a veces resultaron de este tipo de monitoreo, por ejemplo, cambios en la estructura de su sitio Web como resultado de quejas sobre dificultades para encontrar cierta información. En general, ninguna de las organizaciones utilizó estratégicamente los métodos de extracción de datos de las redes sociales. Estas entrevistas se llevaron a cabo, después de todo, en un momento en que el sector público apenas se estaba dando cuenta de la minería de datos. El monitoreo de las redes sociales ocurrió en todas las organizaciones, pero tendió a ser informal y no sistemático. Algunos vieron una brecha aquí, creyendo que la minería de datos en las redes sociales permitiría a las organizaciones detectar y responder a los problemas y preocupaciones locales antes de que se intensificaran, mientras que otros parecían menos convencidos de esta necesidad.

Al plantear inquietudes acerca de las reglas de participación, algunos entrevistados sintieron que no había necesidad de "pescar debates o conversaciones" y expresaron ansiedad sobre "profundizar en el espacio de otra persona". Estas opiniones mostraron que algunos de nuestro pequeño grupo de entrevistados fueron cautelosos con respecto a lo que es y no es aceptable para minar y monitorear en las redes sociales. No obstante, todas las organizaciones identificaron más actividades de minería de datos en las redes sociales en las que podían participar. A través de nuestras entrevistas iniciales que tuvieron lugar durante la investigación, identificamos estas formas en que los participantes imaginaban que la minería de datos podría ser utilizada por sus organizaciones en el futuro:

- Medir el compromiso
- Identificar influyentes clave
- Analizar comentarios
- Monitorear la publicidad
- Identificar lo que le importa al público.
- Políticas de co-diseño.

En relación con el primer punto, al medir la conexión y el compromiso del público con la organización, algunos participantes sintieron que la minería de datos en las redes sociales podría usarse para medir la circulación y el alcance de los mensajes de la organización en las redes locales y para explorar el nivel de interés público y compromiso generado. Otro uso potencial de la minería de datos en las redes sociales fue identificar a los principales influyentes, intermediarios y redes con los cuales interactuar. Los métodos de minería de datos podrían permitir a las organizaciones identificar redes locales de las que no están al tanto, "descubriendo", como lo expresó un entrevistado, "personas que hablan sobre temas que no conocemos" (elecciones, igualdad y participación del Consejo Oficial). Tener una mejor comprensión de la composición de las redes locales podría ayudar a las organizaciones a difundir mensajes

de manera más amplia y efectiva, incluso a grupos "difíciles de alcanzar" a los que no se puede acceder a través de canales convencionales.

La minería de datos en las redes sociales también podría usarse para administrar y analizar consultas y comentarios del público sobre servicios e iniciativas, alertando a las organizaciones sobre problemas particulares que surjan, mientras que la agregación de datos podría usarse para detectar tendencias, positivas y negativas. Del mismo modo, podrían usarse para monitorear la publicidad relevante, o lo que el público, los principales influyentes e intermediarios pueden decir sobre las organizaciones, no a través de comentarios directos sino en otros canales.

Los entrevistados describieron cómo la minería de datos de las redes sociales podría usarse como parte de una estrategia proactiva de gestión de la reputación, en la que la organización extrae las redes sociales para capturar y publicar sentimientos y comentarios positivos sobre la organización. También se puede utilizar como parte de una estrategia más defensiva, donde las organizaciones buscan gestionar los riesgos de reputación. Otros propusieron que los métodos de minería de datos podrían usarse para identificar y analizar lo que el público está diciendo sobre asuntos de interés local, capturando conversaciones públicas y puntos de vista que de otra manera no podrían llegar a la organización a través de canales convencionales, para informar la toma de decisiones de la organización. Un entrevistado, un oficial de elecciones, igualdad y participación de Consejo dijo:

¿Hay alguna manera de capturar esa información y agregarla a la mezcla de que esto es lo que piensa la gente de la ciudad? Sí, no necesariamente puede capturarlo todo, y no puede capturar conversaciones que las personas estén teniendo cara a cara y que no conozca; pero puedes intentar capturar un poco más.

De manera más ambiciosa, algunos participantes descartaron la posibilidad de que los métodos de minería de datos de las redes sociales se puedan utilizar para involucrar al público en el "co-diseño" de políticas. Esto implicaría pasar de un modelo de compromiso de arriba hacia abajo a un enfoque más participativo, donde la organización involucra al público en la formación de políticas. Un entrevistado dijo:

Creo que estamos tratando de avanzar más, en lugar de decir que vamos a presentar esta nueva iniciativa, o estamos pensando en cerrar un edificio, o algo así, necesitamos involucrar a las personas en una etapa temprana en ese co-diseño, obtener la opinión pública sobre qué es lo que quieren de los servicios públicos, cómo quieren que se formen, cómo quieren diseñarlos con nosotros. [...] Herramientas como esta serían realmente útiles para ayudarnos a hacer eso. (Gerente de Inteligencia del Consejo)

Nuestro proyecto de investigación de seis meses, pensado como un experimento sobre lo que era posible con la minería de datos de redes sociales en organizaciones del sector público a pequeña escala y con problemas de liquidez, no iba a poder probar todos los usos potenciales, así que probamos en escala, el potencial de los métodos digitales para

investigar el compromiso público con las organizaciones, así como para examinar redes y conversaciones locales más amplias. En la siguiente sección, analizo algunos experimentos que nuestros socios consideraron beneficiosos. En todos estos ejemplos, destaco el papel desempeñado por un omnipresente "deseo de números".

Comprender el sector público y el deseo de números

Para explorar el compromiso de la gente con nuestras organizaciones asociadas, analizamos el volumen de contribuciones vinculadas a ellas a través del análisis de redes sociales de fuentes de datos particulares, como páginas y grupos de Facebook o cuentas y hashtags de Twitter. Nuestros participantes pensaron que los resultados generados por estos experimentos fueron útiles, ya que el análisis de la red de las páginas institucionales de Facebook o las cuentas de Twitter a través de la herramienta NodeXL resultó especialmente beneficioso. El análisis de red proporcionó información valiosa sobre cuentas activas, páginas y grupos, y el nivel de interacción generado entre los contribuyentes.



Figure 4.1 Networks of participants in relation to 1 month of activity data from the Facebook of one of the partners

La Figura 4.1, por ejemplo, muestra una representación de 1 mes de datos de actividad de una página de Facebook administrada por uno de los socios. Los nodos en el diagrama representan contribuyentes a la página de Facebook y las líneas representan interacciones entre ellos, como las respuestas a las publicaciones. Los contribuyentes se diferencian de acuerdo con su influencia en la red (indicado por el tamaño del nodo) y las comunidades de nodos estrechamente vinculados a los que pertenecen (indicado por el color). Los nombres de los contribuyentes se han eliminado del diagrama para mantener el anonimato a los fines de la publicación, lo que dificulta ver lo que realmente encontramos, pero da una indicación, no obstante, de cómo se representaron

visualmente estos hallazgos. Usando estos métodos, los participantes podrían identificar contribuyentes clave, quién está vinculado a quién, y plataformas particulares podrían compararse en términos del grado de participación pública y actividad que generan.

Utilizamos Gephi, NodeXL y IssueCrawler, así como herramientas comerciales más fáciles de usar para analizar redes locales, y esto también generó datos e ideas relevantes. Usando estos métodos, localizamos grupos importantes de los cuales nuestros socios no estaban al tanto previamente. Por ejemplo, uno de los mapas generados por IssueCrawler (que muestra cómo se comparte y conecta el contenido en línea) identificó una organización de transporte que se estaba comunicando activamente sobre los eventos culturales del grupo del museo y, por lo tanto, podría ser un contacto beneficioso para ellos.

Nuestro análisis también identificó foros locales en línea que estaban activos en términos de contribuciones de los usuarios sobre nuestros socios, pero sobre los cuales, nuevamente, nuestros socios desconocían previamente. Las herramientas comerciales produjeron un análisis de los canales de contribución, indicando la proporción de las contribuciones de cada canal que fue recolectada por cada búsqueda, identificando así los canales clave para cada socio (ver Figura 4.2).

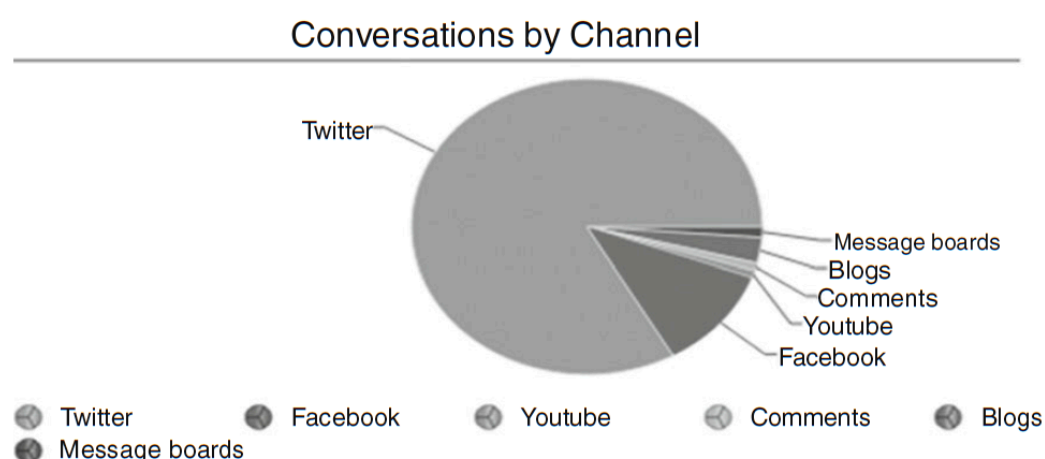


Figure 4.2 Key contribution channels for one particular search, identified by one of the commercial tools

Los datos generados por las herramientas fueron recibidos con entusiasmo por algunos participantes, especialmente cuando se presentan en cuadros y gráficos visualmente atractivos. Los participantes que leyeron los informes que produjimos y los que asistieron al taller expresaron una sensación de asombro. Un participante de la investigación dijo:

Creo que tengo mucha confianza en los números. Me sorprendió lo profundo que pueden llegar muchas de estas herramientas. [...] Creo que son muy inteligentes. Es sorprendente lo mucho que podría profundizar en esto. (Oficial del Equipo Web del Consejo, Departamento de Servicio al Cliente)

En muchos sitios de mi investigación, ese entusiasmo y asombro era común. Caracterizo estas respuestas como indicativas de un deseo generalizado de números, que tiene varios orígenes, no todos los cuales son específicos del fenómeno de volverse ordinario de la minería de datos en las redes sociales. En la década de 1990, Theodore Porter identificó lo que describió como una "confianza generalizada en los números" en un libro con ese nombre (Porter 1995). En *Trust in Numbers*, Porter hizo la pregunta: "¿cómo debemos dar cuenta del prestigio y el poder de los métodos cuantitativos en el mundo moderno?". Argumentó que los números atraen de varias maneras. Son formas familiares y estandarizadas de comunicación y, como tales, se pueden entender desde "lejos", es decir, por personas distanciadas y no familiarizadas con el tema al que se refieren los números. La cuantificación es una tecnología de distancia, argumenta, y Kate Crawford (2013) hace un punto similar sobre Big Data: nos permite ver fenómenos desde lejos, pero esto significa que perdemos los detalles que se pueden observar en un escrutinio más detallado.

Los números son universales, argumenta Porter: se pueden compartir entre culturas, naciones e idiomas. Son impersonales y, por lo tanto, objetivos, y esto a su vez minimiza la necesidad de "confianza personal". Con estas palabras, Porter identifica por qué era necesario cultivar la confianza en los números en un momento particular de la historia marcado por el crecimiento de un estado centralizado y burocrático, administrado por funcionarios "que carecen del mandato de una elección popular". Estos funcionarios necesitaban procesos que parecían objetivos, y esto explica el giro hacia los números y la cuantificación, que no siempre habían recibido la confianza que identifica Porter.

Como señala Porter, "una decisión tomada por números (o por reglas explícitas de algún otro tipo) tiene al menos la apariencia de ser justa e impersonal" (Porter 1996, p. 8). Por lo tanto, el crecimiento de la cuantificación puede verse como un intento de desarrollar medidas de impersonalidad en un momento en que la confianza era escasa; respondió a la incertidumbre en la vida moderna y en las organizaciones modernas: "la objetividad otorga autoridad a los funcionarios", señala Porter (Porter 1996, p. 8). Porter argumenta que esta fe en los números y la credibilidad que se les asigna es "un problema social y moral". Cita la crítica de Horkheimer y Adorno en *Dialectic of Enlightenment* (1948) para corroborar esta afirmación: la ciencia positivista y su dependencia de los números reemplazaron "el concepto con la *fórmula* y la causalidad por regla y *probabilidad*", señalaron (en Porter 1996, p. 18) La cuantificación es un proceso de gestión del mundo, ordenándolo, sin entenderlo: lo que se pierde cuando dominan los números son los entendimientos que las sensibilidades cualitativas nos ayudan a generar, como escribe Baym (2013) en relación con Big Data. En tiempos de datos, la cuantificación está más extendida que nunca, como resultado directo de la ubicuidad de las redes sociales. José van Dijck ofrece una descripción completa y convincente del papel de las plataformas de redes sociales en la dataficación de aspectos de la vida que anteriormente se entendían cualitativamente:

Con el advenimiento de la Web 2.0 y sus sitios de redes sociales en proliferación, se codificaron muchos aspectos de la vida social que nunca antes se habían cuantificado: amistades, intereses, conversaciones casuales, búsquedas de información, expresiones

de gustos, respuestas emocionales, etc. [...] Facebook convirtió actividades sociales como "amistades" y "me gusta" en relaciones algorítmicas (Bucher 2012; Helmond y Gerlitz 2013); Twitter popularizó las personas en línea y promovió ideas creando funciones de "seguidores" y "retuitear" (Kwak et al. 2010); LinkedIn tradujo redes profesionales de empleados y solicitantes de empleo a quién se comunicó con quién, desde qué ubicación y durante cuánto tiempo (van Dijck 2013b); y YouTube *dataficó* el intercambio casual de contenido audiovisual (Ding et al. 2011).

Posteriormente, las interacciones sociales cuantificadas se pusieron a disposición de terceros, ya sean otros usuarios, empresas, agencias gubernamentales u otras plataformas. (van Dijck 2014: 198–9). Además, la retórica que acompaña a la información de este tipo alienta cada vez más la confianza en los números: la afirmación de Anderson, citada anteriormente, de que "con suficientes datos, los números hablan por sí mismos" (2008) ejemplifica esta retórica. Tales afirmaciones existen junto con las condiciones de austeridad y demandas relacionadas con las organizaciones del sector público, y lo que Granovetter llama 'isomorfismo mimético', o la tendencia de las organizaciones a imitarse entre sí ('lo que sea que todos estén haciendo, porque de lo contrario parece no eres una empresa moderna' (1985, citado en Perlin 2011, p. 40)).

No es sorprendente, entonces, encontrar un deseo de números en tales contextos. Pero una confianza en los números y un deseo por los números no son lo mismo. De vuelta en las organizaciones del sector público, algunos participantes fueron más medidos en sus respuestas que los citados anteriormente. A pesar de la afirmación de Anderson, las conclusiones que se pueden observar de los datos extraídos no siempre fueron claras. Un participante explicó que si bien los datos generados por el proyecto proporcionaron información contextual útil, era difícil identificar su valor específico y sacar conclusiones prácticas de él:

Puedo ver que es muy útil ya que proporciona un contexto adicional, pero decir: 'Ah, bueno, con este proyecto, esta es la forma en que debo ir para llegar a donde quiero estar', no hace eso porque en realidad lo que se genera es una visión general en la dirección. (Jefe de Planificación Estratégica, Política y Desempeño, Departamento de Desarrollo de la Ciudad del Consejo)

Otra participante del taller explicó cómo quería ir más allá de la información cuantitativa básica. Ella dijo:

No puedes mirar los datos sobre el valor nominal y decir 'Bueno, él genera muchas menciones y mucho contenido, por lo que es un influyente clave' porque si realmente entras en ese sitio y echas un vistazo a parte del contenido que es generado, no es particularmente de calidad y en realidad no tiene tantos seguidores en ese sitio Web en particular. Continuó diciendo: "Entonces, debes comprender quiénes son las personas, qué están diciendo y cómo eso también es recibido por la audiencia y falta esa parte" (Oficial de Comunicaciones del Consejo).

Con estas palabras, este participante reconoció las limitaciones de los números y la necesidad de datos cualitativos y cuantitativos.