

Factor de Impacto: **2.531**

Factor de impacto de 5 años: **3.183**

☰ Contenido

PDF/ePub

⋮ More

Abstracto

El análisis de eventos de protesta es un método clave para estudiar los movimientos sociales, lo que permite analizar sistemáticamente los eventos de protesta en el tiempo y el espacio. Sin embargo, la codificación manual de eventos de protesta requiere mucho tiempo y recursos. Recientemente, los avances en enfoques automatizados ofrecen oportunidades para codificar múltiples fuentes y crear grandes conjuntos de datos que abarcan muchos países y años. Sin embargo, con demasiada frecuencia los procedimientos utilizados no se discuten en detalle y, por lo tanto, los investigadores tienen una capacidad limitada para evaluar la validez y confiabilidad de los datos. Además, muchos investigadores destacaron los sesgos asociados con el estudio de los eventos de protesta que se informan en las noticias. En este estudio, preguntamos cómo los científicos sociales pueden aprovechar bases de datos de noticias electrónicas y herramientas computacionales para crear datos PEA confiables que cubran una gran cantidad de países durante un largo período de tiempo. Brindamos una descripción detallada de nuestro enfoque semiautomático y ofrecemos una discusión extensa de los posibles sesgos asociados con el estudio de eventos de protesta identificados en fuentes de noticias internacionales.

Introducción

El análisis de la política contenciosa tiene una larga tradición en las ciencias sociales. Uno de los métodos clave en este campo es el análisis de eventos de protesta (PEA), una forma de análisis de contenido que permite recopilar, cuantificar y procesar sistemáticamente grandes cantidades de información sobre protestas políticas a lo largo del tiempo y los países (Hutter, 2014a, [2014b](#)). Los datos de eventos de protesta generalmente se han basado en la cobertura de los periódicos como fuente de datos y, principalmente, en la codificación manual. Sin embargo, la codificación manual de eventos de protesta a través del tiempo y el espacio consume mucho tiempo. Esto plantea un desafío importante para estudiar un gran número de países y cubrir extensos períodos de tiempo. En este artículo, preguntamos cómo los científicos sociales pueden aprovechar bases de datos de noticias electrónicas y herramientas computacionales para crear datos PEA confiables que cubran una gran cantidad de países durante un largo período de tiempo. Responder a esta pregunta de investigación requiere abordar dos desafíos: (a) crear un procedimiento para identificar eventos de protesta únicos en múltiples fuentes de noticias y (b) evaluar la confiabilidad y validez de los datos recopilados.

Siguiendo las sugerencias de [Nardulli et al. \(2015\)](#), desarrollamos procedimientos de varios pasos que combinan la preselección automatizada de documentos relevantes y la codificación manual para obtener un conjunto de datos de alta calidad con un costo y un plazo razonables. [Nardulli et al. \(2015\)](#) proponen automatizar la extracción de contenido fáctico simple del texto y, luego, emplear codificación humana para identificar el contenido que requiere más interpretación. Nuestro enfoque está muy en línea con esta estrategia, aplicamos una línea de herramientas de procesamiento de lenguaje natural (NLP, por sus siglas en inglés) extendida para identificar noticias relevantes antes de la codificación manual de eventos de protesta a partir de ellas. Nuestra decisión de confiar en NLP automatizado para la codificación de los datos requirió la homogeneidad lingüística de nuestras fuentes. Siguiendo el ejemplo de [Beissinger y Sasse \(2014\)](#), elegimos recuperar y codificar eventos de protesta informados por 10 agencias de noticias de habla inglesa, todo lo que pudimos encontrar en los 30 países cubiertos por nuestro estudio. Nuestro enfoque se puede reproducir fácilmente, gracias a nuestra presentación detallada del procedimiento que seguimos.

Además, proporcionamos una evaluación detallada de la fiabilidad y validez de nuestros datos. En cuanto a la parte de codificación manual, hemos alcanzado una aceptable confiabilidad entre codificadores. Además, comparamos nuestros datos con los datos existentes, extraídos de la codificación automática y manual, para proporcionar una evaluación exhaustiva de sus sesgos. Por último, la comparación de nuestros datos con otros datos de eventos de protesta existentes muestra que las fuentes que usamos introducen principalmente los sesgos esperados cuando se usan fuentes internacionales en lugar de nacionales: los eventos incluidos en nuestro conjunto de datos son más dignos de ser noticia para una audiencia internacional, es decir, es más probable que se lleven a cabo en la capital, son eventos más grandes y es más probable que ocurran en momentos en que la comunidad internacional se enfoca en un país determinado. Aparte de este,

El artículo está estructurado de la siguiente manera. Primero, discutimos los enfoques existentes para estudiar eventos de protesta y destacamos los principales límites de los eventos derivados de fuentes de noticias. En segundo lugar, presentamos nuestro procedimiento para la selección automatizada de fuentes de noticias sobre eventos de protesta y la codificación real. Tercero, comparamos nuestros datos con datos de una fuente que utilizó codificación automatizada (Sistema Integrado de Alerta Temprana de Crisis [ICEWS]) y con datos basados en la codificación manual de fuentes de noticias nacionales. En la conclusión, destacamos las fortalezas y debilidades de nuestro enfoque. También discutimos los desafíos futuros para el desarrollo de PEA semiautomático.

Fortalezas, debilidades y avances en PEA

El estudio de los eventos de protesta contribuyó a importantes avances en el campo de la investigación de los movimientos sociales (ver [Earl et al., 2004](#), para una revisión de estas contribuciones). Con el tiempo, las herramientas disponibles para estudiar eventos de protesta evolucionaron, ganando en precisión y volviéndose más reflexivas sobre los sesgos asociados con este método ([Hutter, 2014a](#)). Cualquier método PEA captura solo un segmento de los eventos que tienen lugar en un lugar y tiempo determinado ([Ortiz et al., 2005](#)). Por lo tanto, nunca puede reflejar completamente el flujo de eventos de protesta en la vida real. La mayoría de las investigaciones empíricas en este campo se basan en fuentes periodísticas para estudiar eventos de protesta, por lo que los investigadores introducen sesgos asociados con la cobertura mediática de eventos de protesta en su investigación. Nuestra comprensión de estos sesgos ha avanzado gracias al trabajo crítico en

([Amenta et al., 2009](#) ; [Earl et al., 2004](#) ; [Ortiz et al., 2005](#) ; [Rafail et al., 2019](#)). [Ortiz et al. \(2005\)](#) señalan cuatro tipos de sesgo: características del evento, factores contextuales, estructura de los medios y proceso de investigación. Los primeros tres tipos corresponden a sesgos de selección como el tamaño e intensidad del evento, su proximidad a una agencia de noticias, el ciclo de atención del tema, los motivos de lucro y los intereses corporativos de los medios y sus patrocinadores (ver Earl et al., [2004](#) ; [Ortiz et al., 2005](#) , para una discusión detallada). La buena noticia es que estos sesgos tienden a ser estables en el tiempo y que, cuando los eventos se informan en las noticias, hay pocos errores fácticos (o sesgos descriptivos), solo falta información sobre los objetivos de los movimientos (Earl et al., 2004)).

Entre las características de los eventos que contribuyen a su aparición en las noticias, la proximidad de los medios es un buen predictor de la inclusión de eventos de protesta en las noticias ([Amenta et al., 2017](#) , [Earl et al., 2004](#)). Esto significa que cuanto más cerca esté un evento de un medio de comunicación, es más probable que el evento aparezca en la prensa. Sin embargo, esta proximidad geográfica depende de la intensidad de la movilización ([Rafail et al., 2019](#)). Estos autores identifican un efecto de saturación, la mayor movilización da como resultado la atención de los medios hasta el punto de saturación cuando ya no hay espacio para eventos de protesta adicionales en las noticias.

En los últimos años, el desarrollo de fuentes electrónicas y herramientas de PNL ha facilitado el uso de múltiples fuentes ([Jenkins & Maher, 2016](#)). Algunos son críticos de que construir a partir de más fuentes mejorará la calidad de los datos ([Ortiz et al., 2005](#)), sin embargo, creemos que resuelve algunos problemas asociados con el sesgo de selección de periódicos. Un enfoque de múltiples fuentes permite (a) aumentar la proximidad entre los medios y la protesta cuando se trabaja en varios países; (b) evitar la saturación de noticias cuando la protesta es generalizada en todos los países; y (c) incluyen medios que persiguen diferentes intereses políticos. Sin embargo, esto requiere ser consciente de los sesgos en conflicto entre las fuentes ([Earl et al., 2004](#)) y desarrollar métodos para identificar eventos duplicados para evitar contar eventos únicos varias veces y, por lo tanto, contribuir a una tergiversación de las olas de protesta.

De hecho, los problemas asociados con la identificación duplicada de eventos únicos afectan a los métodos PEA que se basan en herramientas automatizadas ([Schrodt, 2012](#) ; [Wang et al., 2016](#)). Estas nuevas herramientas ofrecen la posibilidad de monitorear eventos de protesta (y conflicto) durante largos períodos de tiempo y espacio. Sin embargo, no cumplen algunas de sus promesas cuando se trata de proporcionar datos fiables y válidos sobre protestas (y conflictos) en tiempo real. Los estudios muestran que estas herramientas tienden a descontextualizar los eventos (centrándose en oraciones individuales en un documento de texto). Por lo tanto, no reconocen si, cuándo y dónde tuvo lugar el evento. [Wang et al. \(2016\)](#), trabajando en un subconjunto seleccionado de eventos de GDELT, encuentran que solo uno de cada dos eventos incluía informes sobre un evento de protesta de la vida real. Además, encuentran que ICEWS es más fuerte para identificar eventos de protesta, pero aún presenta alrededor del 20% de eventos duplicados. Además, estas nuevas herramientas adolecen de una falta de transparencia sobre sus componentes integrados. Es difícil para los investigadores evaluar los procedimientos utilizados para seleccionar eventos y, por lo tanto, evaluar posibles sesgos.

A la luz de estos debates sobre la PEA, nos preguntamos cómo los científicos sociales pueden aprovechar las bases de datos de noticias electrónicas y las herramientas computacionales para crear datos confiables de la PEA de múltiples fuentes que cubran una gran cantidad de países durante largos períodos de tiempo.

Siguiendo a [Zhang y Pan \(2019\)](#) , brindamos información detallada sobre el procedimiento que de

para que otros puedan usarlo para expandir nuestro enfoque semiautomático para PEA. Además, desarrollamos un conjunto de pruebas para evaluar y discutir posibles sesgos en nuestros datos que informan a los investigadores que trabajarán con estos datos y, en general, a aquellos que usan PEA.

Desarrollo de un enfoque semiautomático de varios pasos para PEA

Nuestro objetivo es introducir un procedimiento que permita recuperar información sobre protestas a través de múltiples fuentes y estudiar eventos de protesta que tienen lugar en múltiples países, en nuestro caso 30, durante un largo período de tiempo, 16 años. Para hacer eso, derivamos eventos de protesta de agencias de noticias que publican en inglés y desarrollamos un enfoque de varios pasos basado en herramientas NLP para pasar de los cinco millones de documentos recuperados en el archivo electrónico de noticias a una cantidad manejable de documentos para implementar la codificación manual. . [tabla 1](#) resume nuestro enfoque, que consiste en una serie de pasos de selección para reducir la cantidad de documentos que se codificarán manualmente, reduciéndola a menos del 1% de la cantidad total de documentos identificados por primera vez. Esta tabla destaca la drástica disminución en la cantidad de documentos cuando se pasa del universo de informes noticiosos identificados a través de la búsqueda de palabras clave en múltiples fuentes al universo de documentos que contienen solo aquellos documentos que tienen más probabilidades de informar sobre eventos de protesta. A pesar de esta drástica reducción, la codificación manual requirió, en total, 642 días hábiles para una sola persona. Contamos con un equipo de 35 estudiantes de posgrado para realizar la tarea en un período de 6 meses. Esto ya destaca uno de los principales desafíos para un enfoque de múltiples fuentes, la gran cantidad de datos para procesar. Siguiendo [a Nardulli et al. \(2015\)](#), argumentamos que el uso de herramientas NLP permite reducir la cantidad de documentos que los codificadores humanos necesitan procesar mientras se mantienen los estándares de codificación manual para construir el conjunto de datos final.

Tabla 1 . Pipeline NPL para la Selección de Documentos y Muestreo para Codificación Manual de Eventos de Protesta.						
	Búsqueda por palabra clave	Identificador único de Lexis-nexus	Filtro basado en la ubicación	Clasificador de documentos (doc > .61)	Eliminación casi duplicada	Clasificador de activación de eventos (doc > .85)
Número de documentos	5,251,894	4,211,759	1,116,337	157,572	147,846	101,877
Porcentaje de documentos eliminados	—	19.8	73.5	85.9	6.2	31.1
Porcentaje del total de documentos	100	80.2	21.3	3.0	2.8	1.9

Privacidad

nota _ PNL = procesamiento del lenguaje natural.

- ^a En este paso no eliminamos ningún documento, pero muestreamos los documentos de acuerdo al tamaño y cobertura noticiosa del país en las agencias de noticias internacionales, aplicamos tres tasas de muestreo 0.25, 0.50 y 1 dependiendo de estos criterios. ^b En este paso, la información relevante es la proporción de documentos sin eventos de protesta (no los documentos eliminados). Encontramos un 59,0% de documentos sin eventos.

Selección semiautomática de eventos de protesta: un enfoque de PNL

El primer paso consiste en la extracción de noticias relevantes. Obtenemos informes de noticias del servicio de datos LexisNexis. ¹ Usamos 10 agencias de noticias en inglés ² y recuperamos informes de noticias publicados en el período de 16 años, de 2000 a 2015. Nuestro objetivo es incluir tantas agencias de noticias como sea posible para cubrir con la mayor precisión posible las diferentes regiones y países incluidos en nuestro estudio. De esta forma, seguimos el enfoque multifuente de PEA ([Jenkins & Maher, 2016](#)). La consulta de búsqueda comprende alrededor de 40 palabras clave que describen diferentes acciones de protesta. ³ Optamos por una consulta poco restringida: recupera tantos documentos relevantes como sea posible, incluso a expensas de obtener una gran cantidad de documentos irrelevantes que deben filtrarse en etapas posteriores. Esta búsqueda devuelve un conjunto inicial de 5,2 millones de informes de noticias.

Entre estos 5,2 millones de documentos, terminamos con solo alrededor del 2% de los documentos que informan sobre eventos de protesta. Este número es comparable, si no superior, al que aparece en otros estudios similares ([Oliver, 2019](#)). El primer desafío es identificar y eliminar los documentos irrelevantes. Este es un gran desafío tanto para los procedimientos automatizados como para la codificación manual. Para la codificación automatizada, la dificultad radica en el desequilibrio en el número de eventos en comparación con el número de no eventos, lo que dificulta identificar patrones recurrentes que permitan aprender cómo se presentan los eventos de protesta. Para la codificación manual, la gran cantidad de documentos irrelevantes obliga a los codificadores humanos a leer muchos documentos que no contienen eventos para identificar los pocos eventos informados. Para pasar de tal cantidad de documentos a un conjunto de documentos que se pueden codificar manualmente con un despliegue razonable de recursos, realizamos los siguientes pasos ⁴ :

- Eliminamos los documentos duplicados.
- Utilizamos un filtro de metadatos para descartar documentos no asociados con ninguno de nuestros países de interés.
- Implementamos un clasificador de documentos supervisados y un clasificador de menciones de protestas supervisadas que, en conjunto, intentan distinguir los documentos relevantes de los irrelevantes.

Los duplicados completos representan el 20 % de todos los documentos recuperados y se pueden eliminar rápidamente según los identificadores de documentos. Sin embargo, también nos enfrentamos al problema de informes muy similares que solo difieren en la redacción de una o dos frases. Una gran parte de estos *casi duplicados* se debe a que las agencias de noticias reescriben la misma noticia en el transcurso del día a medida que se desarrollan los eventos. Además, las agencias de noticias a veces comparten noticias entre ellas. Para eliminar casi duplicados, comparamos todos los documentos con el algoritmo SpotSigs ([Theobald et al., 2008](#)).

Privacidad

SpotSigs detecta casi duplicados en dos pasos. En primer lugar, las palabras ancla, en nuestro caso, los términos de búsqueda de la recuperación del documento, se combinan con cadenas cortas de términos de contenido adyacentes para crear firmas de documentos sólidas. En segundo lugar, los documentos se comparan sobre la base de una similitud Jaccard (múltiples) conjuntos entre las firmas de los documentos. ⁵ Identificamos grupos de reportajes con una similitud de Jaccard de 0,75 o superior y mantenemos solo un reportaje de cada grupo en nuestro corpus. ⁶ Nuestro corpus contiene alrededor del 6% de estos documentos casi duplicados. Sin embargo, estos dos pasos de ninguna manera resuelven el problema de los duplicados. Todavía es posible que las mismas o diferentes agencias de noticias informen sobre el mismo evento en documentos redactados de manera muy diferente en diferentes momentos. Así, los hechos destacados se informan en diferentes contextos, por ejemplo, cuando ocurren y cuando sus consecuencias se debaten en los tribunales o en el parlamento o cuando el gobierno reacciona ante ellos o cuando se conmemoran en ocasiones posteriores.

Además del problema de los duplicados, *la ubicación* también plantea un problema. Los informes de noticias pueden ser irrelevantes en términos del alcance geográfico de nuestro análisis. La pregunta de si una historia es relevante para cualquiera de nuestros 30 países es fácil de responder, ya que prácticamente todos los informes de noticias de nuestro corpus vienen con metadatos que también incluyen una lista de países asociados con el contenido del informe. Filtramos los documentos que no incluyen ninguno de nuestros países de interés. Para los países para los que ya tenemos una alta tasa de recuperación (por ejemplo, el Reino Unido, Francia y Alemania), también descartamos documentos en los que este país no se encuentra entre los primeros clasificados y en los que no se encontró ningún otro de nuestros 30 países. . Por lo tanto, nuestro objetivo es aumentar la ponderación de la proporción de documentos de muestra para países donde solo hay unos pocos informes de noticias disponibles ⁷.

Este filtrado basado en la ubicación es el paso más importante para reducir la cantidad de documentos irrelevantes, elimina todos los documentos extraídos de países fuera de Europa ⁸ con aproximadamente tres cuartas partes de los documentos eliminados (73,5 %). La mayor parte de la reducción se debe a la eliminación de documentos que no incluyen ninguno de nuestros 30 países de interés. Solo el 3,5% de los documentos se descartan debido al procedimiento de ponderación. No obstante, la ponderación tiene el efecto deseado, ya que aumenta ligeramente la proporción de documentos de países más pequeños.

A continuación, aplicamos una combinación de un *clasificador de documentos* y un *clasificador activador de eventos* para eliminar aún más los documentos irrelevantes. ⁹ El clasificador de documentos está entrenado en un conjunto de 7.500 documentos, que han sido clasificados manualmente en relevantes (es decir, que mencionan eventos de protesta recientes en Europa) e irrelevantes. No podemos entrar aquí en todos los detalles de este modelo de clasificación. Baste decir que el modelo sigue un enfoque clásico de bolsa de palabras, lo que significa que consideramos todas las palabras en el documento independientemente de las estructuras de las oraciones y las reglas gramaticales, e incluye una serie de pasos que se ha demostrado empíricamente que producen el representaciones de bolsa de palabras más informativas ([Sebastiani, 2002](#)). El modelo de clasificación identifica correctamente el 94% de todos los documentos irrelevantes. Como esto equivale a cientos de miles de documentos debido al desequilibrio de clases, el modelo es una parte importante de nuestro procedimiento de filtrado.

Combinamos este modelo con un enfoque más detallado, un clasificador activador de eventos que nos permite mejorar aún más la identificación de documentos relevantes. Por lo tanto, el supuesto de la bolsa de palabras del modelo de clasificación de documentos no aborda adecuadamente la diversa redacción de los eventos de protesta. Considere las dos oraciones, "los trabajadores se declaran en huelga" y "golpea un gol", cuando tenga en cuenta todas las palabras en un documento, independientemente de las estructuras de las oraciones (ese es el enfoque de la bolsa de palabras), no podemos capturar estos matices. Por lo tanto, ajustamos otro modelo de clasificación para la identificación de menciones de eventos en el texto. Como datos de entrenamiento, usamos 518 documentos que los politólogos anotaron a nivel de palabra para menciones de eventos de protesta. Pedimos a los politólogos que marcaran la extensión de las palabras que, según su juicio, más directamente denotan eventos de protesta. Esto significa que identificamos no solo palabras individuales que se refieren a la protesta, sino también combinaciones específicas de palabras comúnmente utilizadas para referirse a eventos de protesta; por ejemplo, palabras como "salir a la calle", "manifestación violenta", "firmar una petición" y "llevar pancartas". Con este modelo, nuestro objetivo es reducir la identificación automatizada de documentos relevantes para protestas al problema de predecir si un tramo de palabras se refiere a un evento de protesta. En la comunidad de PNL, una palabra como "lanzar" en "arrojar piedras" a veces se denomina desencadenante de eventos (" "firmó una petición" y "llevó pancartas". Con este modelo, nuestro objetivo es reducir la identificación automatizada de documentos relevantes para protestas al problema de predecir si un tramo de palabras se refiere a un evento de protesta. En la comunidad de PNL, una palabra como "lanzar" en "arrojar piedras" a veces se denomina desencadenante de eventos (" "firmó una petición" y "llevó pancartas". Con este modelo, nuestro objetivo es reducir la identificación automatizada de documentos relevantes para protestas al problema de predecir si un tramo de palabras se refiere a un evento de protesta. En la comunidad de PNL, una palabra como "lanzar" en "arrojar piedras" a veces se denomina desencadenante de eventos ([AS, 2005](#)). El modelo eventualmente predice, para un sustantivo común (como "demostraciones") o un verbo (como "lanzar"), si gobierna sintácticamente una probable frase de evento de protesta (como "arrojar piedras"). Una vez más, tampoco podemos entrar aquí en los detalles del enfoque. Basta con señalar que es fácil usar este modelo para la identificación de documentos relevantes para protestas: si el modelo de clasificación no encuentra ningún desencadenante de eventos en un documento, entonces este documento se filtra. Como el modelo asocia estimaciones de probabilidad con todos los factores desencadenantes que identifica en un documento, tomamos la probabilidad más alta que el modelo asigna a uno de los factores desencadenantes como su estimación de la relevancia de ese documento.

Para afinar los dos modelos, experimentamos con la combinación (conjunto) de los dos. Para este propósito, utilizamos un conjunto de aproximadamente 14,000 documentos, que el modelo de clasificación de documentos identificó como relevantes y luego verificó manualmente. Eventualmente, definimos un documento como relevante si ambos modelos de clasificación lo encuentran relevante, de lo contrario, el documento se declara irrelevante. Ajustamos la combinación de manera que logre una mayor precisión y recuperación en este conjunto de datos que el modelo de clasificación de documentos solo. Mantenemos esta configuración para el filtrado final de documentos para la anotación manual. El modelo de clasificación de documentos filtra el 86 % de los documentos de entrada y el modelo de clasificación de activación de eventos otro 30 % de los documentos restantes. Después de aplicar los dos clasificadores, nos queda justo por encima de 100,

Una prueba manual de los dos clasificadores en 1000 documentos muestra que solo hay muy pocos documentos relevantes entre los filtrados por los modelos de clasificador, y esto se aplica a diferentes países y formas de acción. En otras palabras, la proporción de falsos negativos es consistentemente baja en todas las circunstancias evaluadas (2,8 % en promedio). Se ha demostrado que el modelo de activación de eventos funciona mejor para algunos países (como Hungría o el Reino Unido) que para otros (España). Además, el problema inverso de los falsos positivos sigue siendo importante. Aunque la proporción de documentos irrelevantes cae muy rápidamente más allá del umbral de selección, sigue siendo considerable.

Codificación Manual de Eventos de Protesta

En la codificación manual, el objetivo es recuperar información sobre todas las acciones no convencionales motivadas políticamente en los países seleccionados y el período de tiempo. Con este fin, empleamos una versión simplificada del enfoque PEA establecido por [Kriesi et al. \(1995\)](#) y más elaborado por [Hutter \(2014, 2014a\)](#). Dado que el corpus de documentos relevantes resultó ser demasiado grande para administrarlo mediante la codificación manual, tuvimos que reducirlo aún más mediante un procedimiento de muestreo aleatorio. Por lo tanto, para los países para los que seleccionamos una gran muestra de documentos, decidimos codificar manualmente solo el 25 % de los documentos, ¹⁰ para los países con un número medio de documentos, muestreamos el 50 % de todos los documentos para la codificación manual, ¹¹ y para países muy pequeños con solo unos pocos cientos de informes de noticias, todos los informes se codificaron manualmente. ¹²

El primer paso del procedimiento de codificación manual consiste en la identificación de eventos relevantes. Para la identificación de eventos relevantes, nuestros codificadores humanos no se basaron en una definición teórica de acciones de protesta relevantes, que podría ser conceptualmente precisa pero prácticamente muy difícil de implementar. En cambio, los codificadores identificaron eventos relevantes basados en una lista detallada de formas de acción política no convencionales o no institucionalizadas. Pedimos a los codificadores que seleccionaran todas las menciones de acciones de protesta enumeradas en la descripción de la [Tabla 2](#) tan relevante. La definición de eventos en los documentos de los cables de noticias resultó ser una tarea bastante difícil. En primer lugar, un cable de noticias puede contener un resumen de diferentes tipos de eventos políticos y el evento de protesta relevante se menciona, entre otras cosas, y es fácil pasarlo por alto. En segundo lugar, un documento relevante puede contener referencias a más de un evento de protesta, muchas veces puede referirse a una serie de eventos en el mismo lugar (en este caso, la misma ciudad), que pueden ser difíciles de separar entre sí, por ejemplo, una manifestación y un violento enfrentamiento con la policía que involucra a un subgrupo de los manifestantes, una barricada y un teatro de calle. El cable de noticias también puede informar sobre diferentes eventos contemporáneos en diferentes ciudades del mismo país, por ejemplo, manifestaciones en la capital, pero también en varios centros menores en todo el país. Si bien establecimos reglas claras sobre cómo lidiar con tales situaciones, resultó ser difícil aplicarlas en la práctica. En tercer lugar, se puede hacer referencia al mismo evento de protesta varias veces en el documento, pero puede que no siempre sea fácil detectar que en realidad se hace referencia al mismo evento. En cuarto lugar, un evento determinado puede informarse de pasada en un cable de noticias que se centra en otro evento en un país diferente.

Tabla 2 . Formulario de acción y categorías de problemas utilizados en la codificación manual.

Formulario de acción	Descripción
Huelgas	Acción industrial de cualquier tipo (incluidos paros laborales, piquetes)
Demostraciones	Manifestaciones, marchas, mítines, campamentos, reuniones, vigilias y otras reuniones
Peticiones y actividades relacionadas	Peticiones, cartas, activismo de consumidores, boicots, protestas simbólicas (actuaciones, etc.)
Protestas violentas	Sabotaje, disturbios, destrucción de edificios privados o públicos, atentados con bombas o incendios provocados, violencia contra las personas, enfrentamientos con la policía, ciberataques
Otras protestas	Todas las demás formas de acción
Asunto	Descripción
Economía (privado)	Reclamaciones económicas dirigidas a empresas/empleadores: despido de personal, cierre de empresa/sucursal, conflicto laboral relacionado con aumento de sueldo, reducción de sueldo, etc.
Economía (pública)	Reclamaciones económicas dirigidas a instituciones públicas, por ejemplo, bienestar, políticas presupuestarias, políticas agrícolas, regulación laboral, regulación estatal de la economía en general.
Ambiente	Protección del medio ambiente, energía nuclear, otras formas de producción de energía, proyectos de infraestructura (p. ej., transporte), derechos de los animales
Liberalismo cultural	Paz (cuestiones de guerra y paz, armas nucleares y otras convencionales, infraestructura militar, gastos militares, servicio militar), Derechos de las mujeres (incl. igualdad de trato, aborto), LGTB (derechos y reconocimiento de lesbianas, gays, transexuales, bisexuales), Solidaridad internacional (ayuda al desarrollo; antiimperialismo), antirracismo, derechos de los migrantes en general, movilización de ocupantes ilegales (para espacios culturales y de vida autónomos)
Regionalismo	Separatismo, independencia regional, protección de intereses regionales, contramovilización antirregionalista
conservadurismo cultural	Contramovilización a todos los aspectos del liberalismo cultural excepto el antirracismo y los derechos de los migrantes (que está en la xenofobia)
Xenofobia	Extremismo de derecha, movilización racista (contra extranjeros o minorías étnicas), antiinmigración
Político	Representación, corrupción, reformas electorales y temas institucionales en general
Otros	Todos los demás temas no cubiertos por las categorías anteriores

Nuestra definición de un evento es que es único en términos de la combinación de la forma de acción, su momento y su ubicación a nivel de ciudad. Por lo tanto, dos descripciones de eventos de protesta en un documento son iguales si ocurren el mismo día, en la misma ciudad y con la misma forma de acción. Si uno o

Privacidad

más de estos elementos clave de los eventos de protesta difieren, se esperaba que los codificadores codificaran múltiples eventos.

A pesar de todos los esfuerzos por filtrar los documentos irrelevantes, terminamos en la fase de codificación manual con el 59,0 % de los documentos irrelevantes. El 26,3% de los documentos fueron falsos positivos que no contenían ningún evento. En otras palabras, solo 33.655 de los 45.680 documentos muestreados incluían alguna información sobre eventos (ver última columna en la [Tabla 1](#)). Además, sin embargo, otro 32,6% de los documentos contenían solo información sobre eventos que ya estaban informados en otros documentos. La mayoría de estos documentos que contienen duplicados de eventos solamente fueron identificados por los codificadores manuales a medida que revisaban los documentos seleccionados (27,3 %), pero una parte sustancial (5,3 %) también consta de duplicados de documentos que contienen solo eventos que ya fueron codificados por diferentes codificadores de otros documentos. Esto confirma que incluso el procedimiento de selección (semi)automático más sofisticado no es capaz de eliminar el problema de los duplicados de eventos de una manera totalmente satisfactoria. Esto apunta a la necesidad de seguir trabajando en el filtrado automatizado.

Una vez que los codificadores humanos separaron los eventos individuales entre sí, codificaron el siguiente número limitado de variables para cada evento: la fecha del evento, la ubicación del evento (si es posible indicada a nivel de las comunidades), la forma de acción, el problema abordado por la protesta, los actores que participan u organizan el evento de protesta, así como el número de participantes (ver [Tabla 2](#), para la lista de categorías de problemas). Hicimos una prueba de la calidad de esta codificación manual con puntajes de acuerdo entre codificadores recopilados justo antes y durante la codificación. Más precisamente, 14 codificadores recibieron los mismos 65 documentos en diferentes momentos durante su codificación. Medimos el nivel de acuerdo tanto para la identificación de eventos como para la codificación de los atributos del evento. Nos enfocamos en los atributos más importantes, que son los actores, los temas y el número de participantes. Para la identificación de los eventos, evaluamos si dos codificadores concuerdan en la fecha, país y forma de acción de los triples de cada evento que identifican en un mismo documento. Dado que esta es una tarea de identificación abierta, en teoría, hay un número infinito de combinaciones posibles de fecha, país y acción. Informamos la puntuación F1 no ponderada que indica la proporción de eventos coincidentes con la suma de eventos coincidentes y no coincidentes. La puntuación F1 promediada sobre todos los pares posibles de los 14 codificadores es 0,60 con una desviación estándar de 0,06. La puntuación F1 bastante baja refleja la dificultad de la tarea de los codificadores para identificar un evento.

En cuanto a los atributos del evento, primero agregamos las coincidencias y las no coincidencias de los diferentes atributos de cada evento que un par de codificadores identifican conjuntamente en el mismo documento. En segundo lugar, reportamos el Kappa de Cohen sobre el número agregado de acuerdos. Para cada atributo, hay un número fijo de clases, lo que implica que nuestra medida de concordancia debe corregirse al azar. El promedio de Kappa de Cohen es 0,57 (desviación estándar de 0,13) para los actores, 0,53 (0,13) para los temas y 0,45 (0,06) para el número de participantes. Se ha destacado que las pautas para la interpretación de los valores Kappa de Cohen son a menudo engañosas ([Gwet, 2014](#)). Sin embargo, los valores de 0,40 a 0,60 suelen definirse como regulares a buenos. Puede sorprender que estos valores no sean mejores, dado el número limitado de variables a codificar y el bajo número de categorías que usamos en

nuestro libro de códigos. Pero esto muestra que la tarea de codificar eventos de protesta a partir de fuentes de noticias es difícil, incluso para los codificadores humanos, y mucho menos para las máquinas.

Validación externa de los datos

Como ya se discutió, es bien sabido que los datos de eventos de protesta están sujetos a sesgos de selección ([Amenta et al., 2017](#) ; [Earl et al., 2004](#) ; [Ortiz et al., 2005](#) ; [Weidmann, 2016](#)). El hecho de que nuestras fuentes sean agencias de noticias que publican en inglés genera sesgos potenciales específicos. Algunos de estos sesgos están relacionados con las características del evento. Por lo tanto, los eventos violentos, los eventos organizados por fuerzas políticas establecidas o los eventos que coinciden con decisiones políticas importantes tienen más valor noticioso y son informados con mayor frecuencia por las fuentes de noticias en general ([Amenta et al., 2009](#) ; [Oliver & Maney, 2000](#)). Otros sesgos están ligados a la organización de las propias agencias de noticias ([Andrews & Caren, 2010](#)) y a la presencia de una agencia de noticias en un país ([Wu, 2000](#)). Por lo tanto, primero evaluamos la presencia de las 10 agencias de noticias que usamos como fuentes para nuestra investigación en los 30 países incluidos en nuestro estudio. [La Tabla 3](#) muestra que, en conjunto, las 10 agencias de noticias tienen oficinas en todos los países excepto Chipre y Luxemburgo. En la mayoría de los países, hay más de una agencia de noticias en funcionamiento (este es el caso en 22 de nuestros 30 países) y en los países más grandes, como Francia, Alemania e Italia, las agencias de noticias están ubicadas en más de una ciudad. Esto demuestra que nuestro enfoque de fuentes múltiples ofrece una cobertura sólida de la mayoría, si no de todos, los países que estudiamos.

Tabla 3 . Ubicación de las Agencias de Noticias en los 30 Países.

País	Ciudad	Agencias de noticias
Austria	Viena	AFP, APA, DPA
Bélgica	Bruselas	AFP, ANSA, AP, DPA
Bulgaria	Sofía	AFP, APA (vía BTA), DPA
Croacia	Zagreb	AFP, APA (a través de HINA)
Chipre un	—	—
República Checa	Praga	APA (a través de CTK), CTK, DPA
Dinamarca	Copenhague	AFP, DPA
Estonia	Tallin	BNS
Finlandia	helsinki	AFP, DPA
Francia	París	AFP, ANSA, AP, BBC, APA (vía AFP), DPA
Francia	Estrasburgo	DPA
Alemania	Berlina	AFP, DPA

Privacidad

País	Ciudad	Agencias de noticias
Alemania	Colonia	BBC, DPA
Alemania	Fráncfort	AFP, APA (a través de DPA y Minds), DPA
Alemania	Hamburgo	APA (vía DPA), DPA
Grecia	Atenas	AFP, DPA
Hungría	budapest	AFP, APA (vía MTI), DPA, MTI
Islandia	Reikiavik	DPA
Irlanda	Dublín	DPA
Italia	Milán	AFP, ANSA
Italia	Roma	AFP, ANSA, AP, APA (vía ANSA), DPA
Letonia	Anticipar	BNS, DPA
Lituania	Vilna	BNS
Luxemburgo a	—	—
Malta	La Valeta	DPA
Países Bajos	Ámsterdam	DPA
Países Bajos	La Haya	AFP
Noruega	Oslo	AFP, DPA
Polonia	Varsovia	AFP, APA (vía PAP), BBC, DPA, PAP
Portugal	Lisboa	AFP, DPA
Rumania	Bucarest	AFP, APA (vía AGERPRES), DPA
Eslovenia	Bratislava	APA (a través de TASR), DPA
Eslovenia	liubliana	APA (vía STA)
España	Barcelona	AFP
España	Madrid	AFP, ANSA, DPA
Suecia	Estocolmo	AFP, DPA
Suiza	Berna	QUÉ (a través de ATS)
Suiza	Zúrich	AFP
Suiza	Ginebra	AFP, DPA
Reino Unido	Londres	AFP, ANSA, AP, BBC, DPA, PA, APA (vía AFX)

a Ninguna agencia de noticias con sede en este país.

Además, para validar externamente la calidad de nuestros datos, los comparamos con otros dos tipos de conjuntos de datos sobre eventos de protesta. Primero, comparamos nuestros datos con los datos del ICEWS ([Boschee et al., 2015](#)). Los datos de eventos de ICEWS "contienen alrededor de 30 millones de 'historias' que se analizan y codifican utilizando técnicas de PNL basadas en gráficos de palabras utilizando una ontología especialmente desarrollada basada en CAMEO" (Ward et al. [2013](#), págs. 1-10). Los datos provienen de una gran cantidad de fuentes de medios que abarcan fuentes internacionales, regionales, nacionales y locales. En comparación con nuestro enfoque semiautomático, la detección de eventos se realiza de forma totalmente automática. Los cables de noticias son procesados por el software comercial BBN ACCENT, que aplica una variedad de tareas de lenguaje natural y búsquedas en diccionarios para el reconocimiento automático de eventos. Entre otros tipos de eventos, los datos de eventos de ICEWS también incluyen información sobre demostraciones para los 30 países y casi todos los años (todos los años desde 2001 hasta 2014 inclusive) cubiertos por nuestro estudio. Para la forma de acción específica de las demostraciones, la comparación con los datos de ICEWS debería permitirnos mostrar en qué medida nuestros propios datos difieren o mejoran un conjunto de datos creado por procedimientos totalmente automatizados.

En segundo lugar, comparamos nuestro conjunto de datos semiautomático con diferentes conjuntos de datos basados en la codificación manual de artículos de periódicos nacionales e informes de noticias para un subconjunto de países del noroeste, este y sur de Europa. Estos conjuntos de datos han sido recopilados por diferentes equipos de investigadores y, por lo tanto, cubren diferentes períodos de tiempo. ¹³ Nos permiten confrontar el número y tipos de hechos que encontramos en los cables de noticias publicados en inglés con los hechos que encontramos en la prensa nacional. Esto nos permite estimar posibles sesgos relacionados con la escasez relativa de nuestros datos. Esta comparación puede establecer qué tipo de eventos llegan a la prensa internacional y qué eventos se excluyen sistemáticamente.

Comparación con los datos de ICEWS

La pregunta más importante es si el nivel general de demostraciones y la dinámica de la serie temporal proporcionada por nuestros datos son similares a los niveles y dinámicas correspondientes proporcionados por ICEWS. [Figura 1](#) permite una comparación de los niveles y tendencias en las dos series temporales en todos los países. El gráfico de la izquierda muestra los números agregados mensuales de protestas, el de la derecha los promedios móviles de 5 meses. Primero, está claro que el nivel de la serie temporal de ICEWS es mucho más alto que el de nuestros datos. Resulta que (no se muestra aquí), las discrepancias mensuales entre las dos series no siguen un patrón sistemático sino que su desarrollo es más bien estacionario. Esto implica que las discrepancias entre los dos conjuntos de datos son independientes de un período de tiempo específico y, por lo tanto, de eventos específicos. Lo más probable es que las discrepancias sean el resultado de un exceso de informes de manifestaciones en los datos de ICEWS. Esto puede demostrarse mediante un análisis cualitativo detallado de las mayores diferencias mensuales específicas de cada país ([Wueest & Lorenzini, 2020](#)). Por lo tanto, los datos de ICEWS contienen eventos que nuestros codificadores humanos habrían clasificado en otras categorías de formas de acción (p. ej., manifestaciones violentas) o que no se considerarían eventos de protesta

(p. ej., violencia policial). Finalmente, también es muy probable que los datos de ICEWS contengan muchos duplicados de eventos, tal como lo hacía nuestro conjunto de datos antes de que lo limpiáramos manualmente.

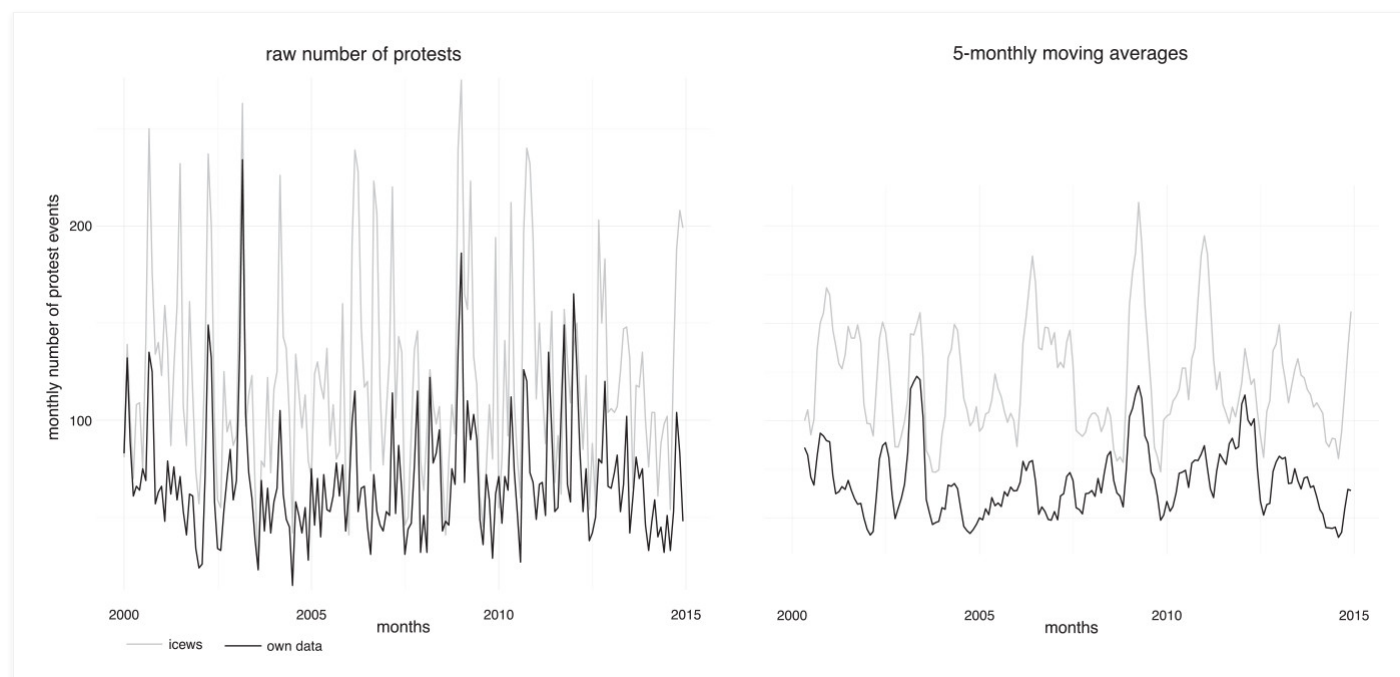


Figura 1 . Número mensual de protestas (izquierda) y promedios móviles de 5 meses (derecha) para ICEWS y nuestros propios datos de protesta.

nota _ ICEWS = Sistema Integrado de Alerta Temprana de Crisis.

En segundo lugar, con respecto al desarrollo a lo largo del tiempo, los datos de ICEWS tienen muchos picos, mientras que nuestros datos solo muestran tres importantes: las manifestaciones de la guerra de Irak en 2003, las protestas posteriores al colapso del sistema financiero mundial en 2008 y las protestas durante la crisis de la eurozona en los años posteriores a 2010. Esto puede deberse a la mayor cantidad y variedad de fuentes utilizadas por ICEWS. En tercer lugar, la comparación de los números sin procesar a la izquierda y los promedios móviles de 5 meses a la derecha muestra que para los análisis descriptivos y visuales, los datos suavizados brindan una mejor idea de lo que está sucediendo.

Las discrepancias entre los dos tipos de datos varían mucho de un país a otro. [Figura 2](#) muestra la distribución de coeficientes de correlación para la serie temporal relacionada con el número de protestas en nuestros 30 países para diferentes niveles de agregación y diversos grados de suavizado. En promedio, obtenemos una mayor congruencia de los dos conjuntos de datos tanto para un mayor nivel de agregación como para un mayor grado de suavizado. Si comparamos series temporales trimestrales en lugar de series mensuales, la mediana de los coeficientes de correlación aumenta de 0,51 a 0,58. Si además aumentamos el grado de suavizado, el coeficiente de correlación de la mediana aumenta hasta el máximo de 0,66 para la serie temporal con medias móviles de 10 trimestres. Al aumentar el grado de suavizado y el nivel de agregación, podemos suponer que algunos de los errores de codificación y selección se corrigen y los conjuntos de datos se vuelven más comparables. Sin embargo, *repartidos* entre los países. La comparación de la serie temporal con las medias

Privacidad

móviles de 10 trimestres es la más extrema. Si bien los dos conjuntos de datos se correlacionan con un impresionante 0,98 para Grecia, la correlación es claramente negativa para Finlandia ($-0,25$). Por tanto, nos enfrentamos a un trade-off entre el nivel medio y la variación de los dos datos. Probando diferentes combinaciones de suavizado y agregación, concluimos que la comparación de las series de tiempo agregadas mensuales suavizadas por un promedio de 5 meses ofrece la mejor solución a esta compensación. La correlación para Finlandia sigue siendo prácticamente inexistente (0,06), pero además de este singular resultado extraño, ningún otro país tiene un coeficiente de correlación inferior a 0,20.

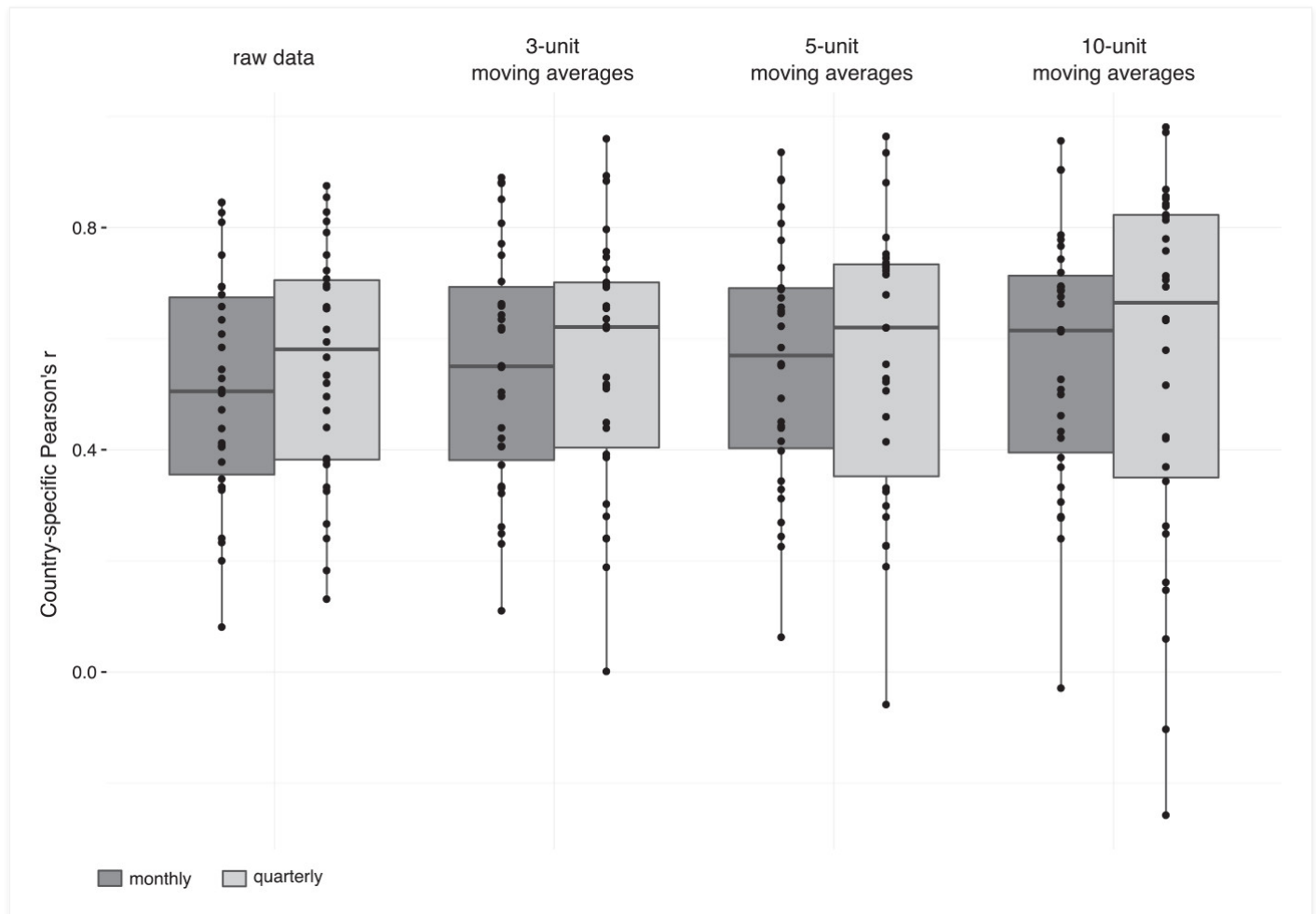


Figura 2 . Correlaciones específicas de cada país entre ICEWS y nuestros propios datos para diferentes niveles de agregación (mensual, trimestral) y diferentes tipos de promedios móviles (3, 5 o 10).

nota _ ICEWS = Sistema Integrado de Alerta Temprana de Crisis.

Análisis más detallados ([Wueest & Lorenzini, 2020](#)) muestran que el tamaño de las diferencias depende de la región. Las diferencias son mayores para el noroeste y el sur de Europa que para el centro y el este de Europa. Esto puede explicarse por el hecho de que las nuevas democracias en Europa Central y Oriental están menos cubiertas por las agencias de noticias internacionales. En consecuencia, el sesgo de selección en las fuentes de ambos conjuntos de datos es menor. Además, el tamaño del país se correlaciona positivamente con las discrepancias entre los dos conjuntos de datos. Dado que las agencias de noticias internacionales

presentes en los países más grandes, la cantidad de eventos de protesta informados en ambos conjuntos de datos es mayor, lo que también implica discrepancias potencialmente mayores. Es probable que las probabilidades de muestreo que aplicamos para nuestra codificación manual por razones de factibilidad también hayan inducido algún sesgo en la generación de datos, por eso ponderamos las probabilidades de muestreo en los análisis de nuestros datos. Finalmente, no encontramos discrepancias sistemáticas entre agencias de noticias específicas o años cubiertos por los dos conjuntos de datos.

Comparación con datos basados en fuentes de noticias nacionales

Para la comparación con las fuentes de noticias nacionales, nos basamos en datos detallados principalmente de tres países, uno en cada región: Hungría, España y Alemania. Estos datos han sido amablemente proporcionados por otros investigadores. Para Hungría, nos basamos en datos de ([Ondřej & Vráblíková, 2019](#)), ¹⁴ de [Martín \(2020\)](#) para España y de [Hutter \(2014b\)](#) para Alemania. Los datos húngaros provienen de la agencia de noticias nacional, mientras que los datos españoles y alemanes provienen de la prensa nacional. [Tabla 4](#) y [5](#) presentamos una comparación de nuestros datos con los datos de las fuentes nacionales con respecto a tres criterios que son particularmente críticos para la cuestión del sesgo: forma de acción, ubicación y tamaño de los eventos de protesta. En general, las discrepancias que encontramos con respecto a estos tres criterios se relacionan en gran medida con la forma en que operan los cables de noticias y con los criterios específicos de interés periodístico que aplican. Es decir, están relacionados con la fuente que utilizamos, la cual, como ya se ha señalado, presenta algunos sesgos bien conocidos.

Tabla 4 . Formas de acción por fuente de noticias para los tres países (porcentajes).

	Alemania		Hungría		España	
	Internacional	Nacional	Internacional	Nacional	Internacional	Nacional
Manifestación.	60.5	48.6	50.0	52.6	44.8	45.0
golpear un	—	—	19.2	5.1	14.1	10.4
Violento	31.2	19.7	19.2	33.9	30.1	19.8
Otro	8.3	31.7	11.6	8.5	11.0	24.8
Total (%)	100	100	100	100	100	100
norte segundo	5,908	1,697	2,104	3,716	5,404	2,002

nota _ Las celdas en negrita tienen residuos ajustados más altos de lo esperado (>1.96). Manifestación. = demostraciones.

^a Los datos de huelga no están disponibles en el conjunto de datos de noticias nacionales de Alemania. ^b El número total de eventos se multiplica por la probabilidad de muestreo. Multiplicamos los eventos encontrados para Alemania, Hungría y España por cuatro.

Tabla 5 . Ubicación y tamaño por fuente de noticias para los tres países (porcentajes).

	Alemania		Hungría		España	
	Internacional	Nacional	Internacional	Nacional	Internacional	Nacional
Capital	23.0	11.6	57.6	36.6	33.9	27.3
Grandes ciudades ^{un}	22.5	33.1	4.2	8.5	16.0	31.1
A nivel nacional ^b	12.0	18.2	28.3	3.3	22.3	0.1
Otro	42.5	37.1	9.9	51.6	27.8	41.5
Total (%)	100	100	100	100	100	100
norte ^c	5,908	1,697	2,104	3,716	5,404	2,002
Grandes eventos (>10.000)	19.7	12.2	17.9	4.8	25.2	16.1
Total ^d (N)	3,576	1,238	896	1,770	2,236	1,363

nota _ Las celdas en negrita tienen residuos ajustados más altos de lo esperado (>1.96).

^a Las grandes ciudades excluyen la capital y comprenden ciudades > 500.000 habitantes para Alemania (Hamburgo, Múnich, Colonia, Fráncfort del Meno, Essen, Stuttgart, Dortmund, Dusseldorf y Bremen), y para España (Barcelona, Valencia, Sevilla, Zaragoza y Málaga).); y >100.000 habitantes para Hungría (Miskolc, Nyiregyhaza, Debrecen, Keckskemet, Győr y Szeged). ^b Para Alemania y España, los datos resultantes de la codificación de las noticias nacionales no incluyen un código específico para los eventos de protesta que tienen lugar en todo el país. Creamos un proxy utilizando un código para identificar eventos que tienen lugar en la misma fecha, abordan los mismos problemas, pero tienen lugar en diferentes lugares. ^cEl número total de eventos se multiplica por la probabilidad de muestreo. Multiplicamos los eventos encontrados para Alemania, Hungría y España por cuatro. ^d El número total de eventos es menor aquí ya que para muchos eventos el número de participantes no se menciona en las noticias.

Primero, con respecto a las formas de acción, encontramos que las manifestaciones representan aproximadamente la mitad de los eventos, independientemente de la fuente; los datos alemanes no son comparables a este respecto, porque no incluyen ninguna huelga. Para Hungría y España, las fuentes internacionales reportan una mayor proporción de huelgas que las nacionales. Con respecto a la acción violenta, observamos diferencias entre fuentes nacionales e internacionales, pero estas varían entre países. Mientras que en Hungría las fuentes nacionales informan de más hechos violentos, en España incluyen menos. Las diferencias específicas de cada país con respecto a la información sobre la violencia están relacionadas con el hecho de que la violencia en Hungría a menudo ocurre en las regiones rurales, algo que las agencias de noticias internacionales pasan por alto, mientras que en España, la violencia la comete principalmente ETA, [La tabla 4](#) también muestra que otras formas de actuación representan una proporción menor del total de acciones reportadas para España en las noticias internacionales (11,0% frente al 24,8% en las noticias

nacionales). De manera similar, en Alemania, las noticias internacionales incluyen menos “otras formas de acción” porque es muy difícil capturar las formas no convencionales de protesta que se agrupan en la categoría residual. Por ejemplo, formas de acción como los activistas ambientales que se sientan en los árboles para evitar que los corten, o los ambientalistas que interceptan barcos balleneros con botes de goma son muy difíciles de capturar sistemáticamente.

La comparación con respecto a la ubicación presentada en [la Tabla 5](#) revela que las fuentes de noticias internacionales exageran las protestas que tienen lugar en la capital de cada uno de los tres países, y en el país en su conjunto en dos de los tres casos (en Hungría y España, pero no en Alemania). Nuevamente, esto es el resultado de la forma en que operan los cables de noticias internacionales: prestan mayor atención a los eventos en la capital de un país y, a menudo, informan de manera resumida sobre eventos que tienen lugar simultáneamente en diferentes lugares del mismo país. Nuevamente, los datos alemanes no son del todo comparables con respecto a la categoría de resumen "país en su conjunto", lo que explica por qué no encontramos una discrepancia correspondiente en este caso. En tercer lugar, la categoría residual, otras ubicaciones, aparece con más frecuencia en los datos extraídos de fuentes nacionales que internacionales. Este es el caso de Hungría y de España. Último, Los cables de noticias también son selectivos con respecto al tamaño de los eventos de protesta medidos por el número de participantes: claramente se enfocan en eventos grandes, mientras que los eventos más pequeños tienden a escapar de ellos. Esto es particularmente cierto en un país pequeño como Hungría sin un servicio de noticias en inglés propio, pero menos en los países más grandes. Así, encontramos el mismo efecto, pero en menor medida para Alemania, pero no para España, donde nuestras agencias de noticias parecen haber sido bastante representativas en cuanto al tamaño de los eventos.

[Tabla 6](#) informa el poder explicativo de diferentes conjuntos de predictores que dan cuenta del valor noticioso de eventos de protesta específicos. Realizamos una regresión logística sobre eventos de protesta informados en las noticias internacionales para identificar los conjuntos de predictores que dan cuenta de la cobertura de protestas en las noticias internacionales en comparación con las noticias nacionales. Presentamos nuestros conjuntos de predictores siguiendo un enfoque gradual, los predictores incluyen las tres características de los eventos de protesta discutidos anteriormente (ubicación, tamaño y forma de acción), así como el año del evento de protesta, el problema abordado y los actores que participan. en el caso. Realizamos la regresión logística por separado para seis países, es decir, los tres países ya discutidos anteriormente (Alemania, Hungría y España) y tres más (Polonia, los Países Bajos y el Reino Unido).

Tabla 6 . Poder explicativo de diferentes conjuntos de predictores entre países (Pseudo R^2 para regresiones logísticas).							
Modelos paso a paso	Conjuntos de predictores agregados al modelo	España	Hungría	Polonia	Los países bajos	Alemania	Gran Bretaña
Paso 1	Ubicación	.148	.184	.043	.090	.026	.046
Paso 2	Tamaño	.197	.200	.102	.107	.045	.063
Paso 3	Formulario de acción	.228	.210	.135	.122	.121	.067

Privacidad

Modelos paso a paso	Conjuntos de predictores agregados al modelo	España	Hungría	Polonia	Los países bajos	Alemania	Gran Bretaña
Paso 3b	Formulario de acción # ubicación		.262				
Etapa 4	Año	.251	.289	.150	.148	.127	.086
Paso 5	Asuntos	.282	.301	.178	.148	.144	.115
modelo completo	Actores	.459	.316	.188	.178	.156	.120
	norte	3,353	4,242	7,555	692	3,174	2,192

nota _ Esta tabla presenta el pseudo R^2 para las regresiones logísticas paso a paso que ejecutamos en seis países. Las variables dependientes que predecimos miden si los eventos de protesta aparecen en las noticias internacionales o nacionales. Las unidades de análisis son eventos de protesta identificados como una forma de acción específica que tiene lugar en una fecha determinada (un día de 24 horas) en un lugar determinado (una ciudad o un país entero). Los diferentes conjuntos de predictores agregados como variables independientes en cada paso de la regresión logística dan cuenta de las características de los eventos de protesta que influyen en su valor periodístico. Como predictores usamos la ubicación (eventos que ocurren en la capital, en otras ciudades grandes y en todo el país), tamaño del evento, formas de acción (manifestaciones, huelgas, evento violento, otros), el año (2000 a 2011), temas (económicos y políticos) y actores (partidos políticos, sindicatos, otras organizaciones de la sociedad civil). En los países occidentales (Alemania, Gran Bretaña y los Países Bajos), los conjuntos de datos nacionales no incluyen huelgas (por lo que solo consideramos tres formas de acción) y no tenemos una variable para capturar cuestiones políticas (por lo que nos enfocamos en cuestiones económicas).).

Mirando primero el poder predictivo del modelo completo, los resultados muestran que el poder predictivo del conjunto de seis predictores es mayor para Hungría y España, dos países que no tienen sus propios servicios de noticias en inglés; en España explicamos hasta el 45,9% de las discrepancias y en Hungría el 31,6%. Por el contrario, el poder explicativo de los predictores es más bajo para Alemania (15,6 %) y Gran Bretaña (12,0 %), dos grandes países con sus propios servicios de noticias en inglés. En estos casos, la discrepancia es más difícil de predecir por nuestro conjunto de predictores, porque es menos probable que dependa de la forma en que operan los cables de noticias internacionales, sino que depende de las idiosincrasias de las fuentes respectivas. Polonia (con un 18,8 % explicado) y los Países Bajos (con un 17,8 % explicado) se sitúan entre los dos tipos de países: en estos dos países,

Al observar conjuntos específicos de variables, observamos que en España y Hungría, la ubicación de los eventos ya tiene un fuerte poder explicativo (14,8% y 18,4%, respectivamente). Esto es exactamente lo que esperábamos en función de la forma en que operan los cables de noticias internacionales, centrándose en eventos que tienen lugar en la capital, en las grandes ciudades o en todo el país. Sin embargo, en los otros cuatro países, la ubicación de los hechos juega un papel más limitado a la hora de explicar el hecho en las noticias internacionales (entre el 9 % en los Países Bajos y el 2,6 % en Alemania). Sumando el tamaño de los eventos, aumenta aún más el poder predictivo en un 4,9% en España y en un 5,9% en Polonia. Por el contrario, en Hungría, el tamaño por sí solo añade poco poder predictivo al modelo. Sin embargo, cuando in

formas de acción, así como la interacción entre las formas de acción y la ubicación, el modelo explica el 26,1% de la varianza entre los informes de noticias nacionales e internacionales. De manera más general, agregar formularios de acción al modelo aporta poco en términos de poder explicativo (entre 0,4% en Gran Bretaña y 3,3% en Polonia), con la única excepción de Alemania, donde los formularios de acción agregan 7,6% al poder explicativo. En este país grande, con más de una agencia de noticias nacional, la forma de acción explica las diferencias en los reportajes cuando se comparan las noticias internacionales con las nacionales, mientras que la ubicación y el tamaño son menos importantes. la única excepción es Alemania, donde las formas de acción añaden un 7,6% al poder explicativo. En este país grande, con más de una agencia de noticias nacional, la forma de acción explica las diferencias en los reportajes cuando se comparan las noticias internacionales con las nacionales, mientras que la ubicación y el tamaño son menos importantes. la única excepción es Alemania, donde las formas de acción añaden un 7,6% al poder explicativo. En este país grande, con más de una agencia de noticias nacional, la forma de acción explica las diferencias en los reportajes cuando se comparan las noticias internacionales con las nacionales, mientras que la ubicación y el tamaño son menos importantes.

La tabla revela variaciones más pequeñas entre países cuando agregamos predictores que representan años, problemas y actores. La única excepción para estos tres conjuntos de predictores se refiere al alto poder predictivo de los actores en España. En este caso, agregar partidos políticos, sindicatos y otras organizaciones de la sociedad civil como predictores aumenta el poder predictivo del modelo en un 17,7%. Esto se relaciona con dos características distintas de la prensa internacional: (a) presta más atención a eventos que no son organizados por partidos políticos, sindicatos y otras organizaciones de la sociedad civil (como el movimiento de indignados y acciones de protesta relacionadas) que a los noticias nacionales; y (b) es menos probable que las noticias internacionales den cuenta de la presencia de estos actores en el evento de protesta. En general,

Conclusión

Hemos descrito nuestro procedimiento para identificar y codificar eventos de protesta en 30 países europeos durante el período 2000-2015. Aplicamos un procedimiento híbrido que combina las fortalezas de las máquinas y los humanos para codificar un amplio conjunto de eventos políticos. Confiamos en una canalización de herramientas de NLP extendida para identificar documentos de noticias relevantes antes de la codificación manual de eventos de protesta a partir de ellos. Si bien podemos estar bastante seguros de que la proporción de documentos relevantes filtrados por nuestro procedimiento de selección es muy baja (cerca del 5 %), debemos tener en cuenta que incluso los porcentajes bajos de falsos negativos aún pueden generar grandes números absolutos, dado que hemos filtrado millones de documentos que se seleccionaron originalmente en función de las palabras clave. Nuestro procedimiento de selección tampoco fue capaz de identificar exclusivamente documentos relevantes, ya que resultó que aproximadamente una cuarta parte (26,3%) de los documentos seleccionados no contenían ningún evento. Además, aproximadamente un tercio de los documentos (32,6 %) contenían solo duplicados de eventos que ya habían sido informados en otros documentos.

En la fase de codificación manual, la tarea era doble: identificar todos los eventos relevantes en un documento dado y luego codificar un número limitado de atributos. Como indican nuestras pruebas de confiabilidad, ambas tareas resultaron difíciles incluso para los codificadores manuales, pero logramos una calidad de codificación razonablemente buena.

Las comparaciones de nuestro conjunto de datos con un conjunto de datos producido por procedimientos completamente automatizados (los datos de ICEWS sobre demostraciones) y con conjuntos de datos producidos por selección y codificación manual (conjuntos de datos seleccionados para varios países de las tres regiones) sugiere que cada tipo de datos tiene sus ventajas y desventajas. Los datos de ICEWS brindan muchas más demostraciones que nuestro conjunto de datos, pero nuestro análisis detallado sugiere que esto se debe, al menos en parte, a un exceso de información por parte de ICEWS. La dinámica a lo largo del tiempo es bastante similar para algunos países (p. ej., Grecia), pero muy diferente para otros (p. ej., Finlandia). Hemos descubierto algunos de los factores responsables de las discrepancias, pero en el análisis final no podemos decir qué tipo de datos es más capaz de reproducir la dinámica "real". En cuanto a las comparaciones con conjuntos de datos producidos por selección y codificación manual, detectamos una serie de discrepancias que están claramente vinculadas a la forma en que operan los cables de noticias internacionales: tienden a prestar especial atención a los grandes eventos que tienen lugar en la capital nacional o "en todo el país", y a eventos durante los períodos pico de protesta en un país determinado. Estos factores explican mejor las discrepancias de los países que no tienen su propia agencia nacional de noticias en inglés. Para los países que tienen tales agencias de noticias, no pudimos dar cuenta de las discrepancias entre nuestros datos y los datos identificados y codificados manualmente en las fuentes de noticias nacionales. Tienden a prestar especial atención a los grandes eventos que tienen lugar en la capital nacional o "en todo el país", y a los eventos durante los períodos de máxima protesta en un país determinado. Estos factores explican mejor las discrepancias de los países que no tienen su propia agencia nacional de noticias en inglés. Para los países que tienen tales agencias de noticias, no pudimos dar cuenta de las discrepancias entre nuestros datos y los datos identificados y codificados manualmente en las fuentes de noticias nacionales. Tienden a prestar especial atención a los grandes eventos que tienen lugar en la capital nacional o "en todo el país", y a los eventos durante los períodos de máxima protesta en un país determinado. Estos factores explican mejor las discrepancias de los países que no tienen su propia agencia nacional de noticias en inglés. Para los países que tienen tales agencias de noticias, no pudimos dar cuenta de las discrepancias entre nuestros datos y los datos identificados y codificados manualmente en las fuentes de noticias nacionales.

Nuestro estudio ofrece una discusión detallada de un enfoque semiautomático para recuperar eventos de protesta de agencias de noticias internacionales. Además, proporciona información clave sobre los sesgos que este método conlleva en el conjunto de datos final. Una nueva investigación en el campo combina el estudio automatizado de texto e imágenes para identificar protestas en las redes sociales ([Zhang & Pan, 2019](#)). Además, se necesita más investigación para combinar fuentes escritas en diferentes idiomas y estudiar las formas de acción simbólica que son más difíciles de identificar tanto para codificadores humanos como para máquinas.

Declaración de Conflicto de Intereses

El(los) autor(es) declaró(n) que no existe(n) ningún conflicto de interés potencial con respecto a la investigación, autoría y/o publicación de este artículo.

Fondos

El (los) autor (es) declararon haber recibido el siguiente apoyo financiero para la investigación, autoría y/o publicación de este artículo: El proyecto fue apoyado por la subvención Nr. 338875 y por la subvención SNF Nr. 100017_146104.

notas al pie

1. <http://www.lexisnexis.com/>

[IR A LA NOTA AL PIE](#)

2. Incluimos las siguientes agencias de noticias: AFP, AP, APA, BBC, BNS, CTK, DPA, MTI, PA y PAP. Nuestro objetivo era incluir las principales agencias de noticias (AFP, DPA y PA) y algunas regionales que cubrían más a fondo Europa del Este y del Sur.

[IR A LA NOTA AL PIE](#)

3. Cadena de consulta “¡iniciativa O referéndum O petición! O firma! ¡O campaña! ¡O protesta! O demostrar! ¡O manifiesto! ¡O marcha! ¡O marcha! ¡O desfile O mitin! O piquete! O (cadena humana) O motín! O refriega O festival O ceremonia O (teatro de calle) O (espectáculo itinerante) O vigilia O huelga! ¡O boicot! ¡O bloquear! ¡O sentarse o ponerse en cuclillas! O motín! ¡O bomba! ¡O bomba incendiaria! ¡O molotov O grafiti O asalto O ataque O incendio provocado O incendiario! O (fuego l/1 aumentando) O (encendido Y en llamas) O mina terrestre O sabot! ¡O rehén! ¡O asesinar! O disparó O asesinó O mató”.

[IR A LA NOTA AL PIE](#)

4. [La Tabla 1](#) muestra el orden en que se ha aplicado la canalización. Los dos pasos de deduplicación no se suceden en esta canalización. Dado que conceptualmente pertenecen juntos, los discutimos juntos en el texto.

[IR A LA NOTA AL PIE](#)

5. La similitud de Jaccard se define como el tamaño de la intersección dividido por el tamaño de la unión de las firmas del documento. Para nuestra implementación, consulte

<https://gitlab.cl.uzh.ch/rothenha/nearDuplicateDetection/>

[IR A LA NOTA AL PIE](#)

6. Elegimos el valor de umbral óptimo evaluando manualmente 100 pares duplicados para cada uno de los siguientes umbrales: 0,65, 0,70, 0,75 y 0,80. Para un valor umbral de 0,75, no encontramos errores.

[IR A LA NOTA AL PIE](#)

7. Estos países son AUT, BEL, BGR, CHE, CYP, DNK, FIN, GRC, ISL, LUX, MLT, NLD, PRT, ROU, SVK, SVN y SWE.

[IR A LA NOTA AL PIE](#)

8. Europa aquí se refiere a EU-27 más Islandia, Noruega y Suiza.

[IR A LA NOTA AL PIE](#)

9. En el aprendizaje automático, un modelo de clasificación (p. ej., regresión logística) se denomina clasificador. Entrenar un clasificador significa ajustar un modelo de clasificación a los datos. Las variables independientes se denominan características ([Justin & Stewart, 2013](#)).

[IR A LA NOTA AL PIE](#)

10. CZE, DEU, ESP, FRA, GBR, HUN, IRL, ITA, LVA y POL.

[IR A LA NOTA AL PIE](#)

11. BEL, DNK, ESTE, GRC, NOR, PRT y SVK.

[IR A LA NOTA AL PIE](#)

12. AUT, BGR, CHE, CYP, FIN, ISL, LTU, LUX, MLT, NLD, ROU, SVN y SWE.

[IR A LA NOTA AL PIE](#)

13. Las fuentes son MTI (Hungarian News Agency Corporation) para Hungría, El País para España, Frankfurter Rundschau para Alemania, PAP (Polish Agency Press) para Polonia, CTK (Czech News Agency) para la República Checa, TASR (Agencia de la República Eslovaca) para Eslovaquia, The Guardian para el Reino Unido y NRC/Handelsblad para los Países Bajos.

[IR A LA NOTA AL PIE](#)

14. El autor proporcionó una actualización de su conjunto de datos para Hungría y Polonia para el período 2011-2015 a pedido nuestro. El estudio original abarcó el período 1989-2011 (ver Ondrej Cisar: Acción colectiva y protesta en Europa Central y Oriental, propuesta de proyecto presentada a la Fundación Checa para la Ciencia —Parte C. Descripción del proyecto, número de registro P404/11/0462).

[IR A LA NOTA AL PIE](#)

Referencias

AS. (2005). *Directrices de anotación en inglés de Ace (Extracción automática de contenido)* . www ldc upenn edu/sites/www ldc upenn edu/files/english-relations-guidelines-v6.2.pdf

[IR A REFERENCIA](#)

[Google Académico](#)

Amenta E., Caren N., Olasky SJ, Stobaugh JE (2009). Todos los movimientos caben para imprimir: Quién, qué, cuándo, dónde y por qué aparecieron las familias smo en el New York Times en el siglo XX. *Revisión Sociológica Americana* , 74(4), 636-656. <https://doi.org/10.1177/000312240907400407>

[+ SHOW REFERENCES](#)[Google Académico](#)

Amenta E., Elliott TA, Shortt N., Tierney AC, Türkoğlu D., Vann B. Jr. (2017). Del sesgo a la cobertura: lo que explica cómo las organizaciones de noticias tratan a los movimientos sociales. *Compás de Sociología* , 11(3), Artículo e12460. <https://doi.org/10.1111/soc4.12460>

[+ SHOW REFERENCES](#)[Google Académico](#)

Andrews KT, Caren N. (2010). Haciendo las noticias: Organizaciones del movimiento, atención de los medios y la agenda pública. *Revisión Sociológica Americana* , 75(6), 841-866. <https://doi.org/10.1177/0003122410386689>

[IR A REFERENCIA](#)[Google Académico](#)

Beissinger MR, Sasse G. (2014). ¿El fin de la "paciencia"? La Gran Recesión y la protesta económica en Europa del Este. En Bermeo N., Bartels LM (Eds.), *Política de masas en tiempos difíciles: opiniones, votos y protestas en la Gran Recesión* (págs. 334-370). Prensa de la Universidad de Oxford.

[IR A REFERENCIA](#)[referencia cruzada](#)[Google Académico](#)

Boschee E., Lautenschlager J., O'Brien S., Shellman S., Starz J., Ward M. (2015). *Datos de eventos codificados por ICEWS* (Versión 30). Dataverso de Harvard. <https://doi.org/10.7910/DVN/28075>

[IR A REFERENCIA](#)[Google Académico](#)

Earl J., Martin A., McCarthy JD, Soule SA (2004). El uso de datos periodísticos en el estudio de la acción colectiva. *Revisión Anual de Sociología* , 30(1), 65-80. <https://doi.org/10.1146/annurev.soc.30.012703.110603>

[+ SHOW REFERENCES](#)[Privacidad](#)

[Google Académico](#)

Gwet KL (2014). *Manual de confiabilidad entre evaluadores: la guía definitiva para medir el grado de acuerdo entre evaluadores*. Analítica avanzada.

[IR A REFERENCIA](#)[Google Académico](#)

Hutter S. (2014a). Análisis de eventos de protesta y su descendencia. En della Porta D. (Ed.), *Prácticas metodológicas en la investigación de movimientos sociales* (págs. 335-367). Prensa de la Universidad de Oxford.

[+ SHOW REFERENCES](#)[referencia cruzada](#)[Google Académico](#)

Hutter S. (2014b). *Protesta por la cultura y la economía en Europa occidental: nuevas divisiones en la política de izquierda y derecha*. Prensa de la Universidad de Minnesota.

[+ SHOW REFERENCES](#)[referencia cruzada](#)[Google Académico](#)

Jenkins JC, Maher TV (2016). ¿Qué debemos hacer con respecto a la selección de fuentes en los datos de eventos? Desafíos, avances y posibles soluciones. *Revista Internacional de Sociología*, 46(1), 42-57.

<https://doi.org/10.1080/00207659.2016.1130419>[+ SHOW REFERENCES](#)[Google Académico](#)

Justin G., Stewart BM (2013). Texto como datos: la promesa y las trampas de los métodos automáticos de análisis de contenido para textos políticos. *Análisis político*, 21(3), 267-297.

[IR A REFERENCIA](#)[referencia cruzada](#)[Google Académico](#)

Kriesi H., Koopmans R., Duyvendak JW, Giugni MG (1995). *Nuevos movimientos sociales en Europa Occidental*. Prensa de la Universidad de Minnesota.

[IR A REFERENCIA](#)[Google Académico](#)

Martín P. (2020). *Quejas y protestas públicas: Movilización política en España en tiempos de austeridad*. Palgrave Macmillan.

[IR A REFERENCIA](#)[Google Académico](#)

Nardulli PF, Althaus SL, Hayes M. (2015). Un enfoque progresivo de aprendizaje supervisado para generar abundantes datos sobre conflictos civiles. *Metodología Sociológica*, 45(1), 148-183.

<https://doi.org/10.1177/0081175015581378>[+ SHOW REFERENCES](#)[Google Académico](#)

Oliver P. (2019). Los grandes métodos revelan sus propias limitaciones. *Metodología Sociológica*, 49(1), 63-68.

<https://doi.org/10.1177/0081175019860236>[IR A REFERENCIA](#)[Google Académico](#)

Oliver PE, Maney GM (2000). Procesos políticos y cobertura de periódicos locales de eventos de protesta: del sesgo de selección a las interacciones triádicas. *Diario americano de sociología*, 106(2), 463-505.

<https://doi.org/10.1086/316964>[IR A REFERENCIA](#)[Google Académico](#)

Ondřej C., Vráblíková K. (2019). Agenda de protesta nacional y la dimensionalidad de la política de partidos: evidencia de cuatro democracias de Europa central y oriental. *Revista europea de investigación política*, 58(4), 1152-1171.

[IR A REFERENCIA](#)[referencia cruzada](#)

[Google Académico](#)

Ortiz D., Myers D., Walls E., Díaz M.-E. (2005). ¿Dónde estamos con los datos de los periódicos? *Movilización: An International Quarterly* , 10(3), 397-419. <https://doi.org/10.17813/maiq.10.3.8360r760k3277t42>

[+ SHOW REFERENCES](#)[Google Académico](#)

Rafael P., McCarthy JD, Sullivan S. (2019). Los climas locales de receptividad y las dinámicas de atención mediática a la protesta. *Movilización: An International Quarterly* , 24(1), 1-18. <https://doi.org/10.17813/1086-671X-24-1-1>

[+ SHOW REFERENCES](#)[Google Académico](#)

Schrodt PA (2012). Precedentes, avances y perspectivas en datos de eventos políticos. *Interacciones internacionales* , 38(4), 546-569. <https://doi.org/10.1080/03050629.2012.697430>

[IR A REFERENCIA](#)[Google Académico](#)

Sebastiani F. (2002). Aprendizaje automático en la categorización automática de texto. *Encuestas informáticas de ACM* , 34(1), 1-47. <https://doi.org/10.1145/505282.505283>

[IR A REFERENCIA](#)[Google Académico](#)

Theobald M., Siddharth J., Paepcke A. (2008). Spotsigs: Detección de casi duplicados robusta y eficiente en grandes colecciones web. En *Actas de la 31.ª Conferencia Internacional Anual ACM SIGIR sobre Investigación y Desarrollo en Recuperación de Información* (págs. 563-570). Asociación para Maquinaria de Computación.

[IR A REFERENCIA](#)[referencia cruzada](#)[Google Académico](#)

Wang W., Kennedy R., Lazer D., Ramakrishnan N. (2016). Dolores de crecimiento para el monitoreo global de eventos sociales. *Ciencia*, 353(6307), 1502-1503. <https://doi.org/10.1126/science.aaf6758>

[+ SHOW REFERENCES](#)

[Google Académico](#)

Ward MD, Beger A., Cutler J., Dickenson M., Dorff C., Radford B. (2013). Comparación de datos de eventos GDELT e ICEWS. *Análisis*, 21(1), 1-10.

[IR A REFERENCIA](#)

[Google Académico](#)

Nota de Weidmann (2016). Una mirada más cercana al sesgo de informe en los datos de eventos de conflicto. *Revista estadounidense de ciencia política*, 60(1), 206-218. <https://doi.org/10.1111/ajps.12196>

[IR A REFERENCIA](#)

[Google Académico](#)

Wu D. (2000). Determinantes sistémicos de la cobertura de noticias internacionales: una comparación de 38 países. *Revista de comunicación*, 50(2), 110-130. <https://doi.org/10.1111/j.1460-2466.2000.tb02844.x>

[IR A REFERENCIA](#)

[Google Académico](#)

Wueest B., Lorenzini J. (2020). Validación externa del análisis de eventos de protesta. En Kriesi H., Lorenzini J., Wüest B., Hausermann S. (Eds.), *Contención en tiempos de crisis: Recesión y protesta política en treinta países europeos* (pp. 49-76). Prensa de la Universidad de Cambridge.

[+ SHOW REFERENCES](#)

[referencia cruzada](#)

[Google Académico](#)

Zhang H., Pan J. (2019). Casm: un enfoque de aprendizaje profundo para identificar eventos de acción colectiva con datos de texto e imágenes de las redes sociales. *Metodología Sociológica*, 49(1), 1-57. <https://doi.org/10.1177/0081175019860244>

[+ SHOW REFERENCES](#)

[Google Académico](#)

Biografías

Jasmine Lorenzini es investigadora del Instituto de Estudios de Ciudadanía de la Universidad de Ginebra.

Hanspeter Kriesi es profesor en el Departamento de Ciencias Políticas y Sociales del Instituto Universitario Europeo de Florencia.

Peter Makarov es investigador de doctorado en el Instituto de Lingüística Computacional de la Universidad de Zúrich.

Bruno Wüest es investigador en Sotomo en Zúrich. Fue investigador posdoctoral en el Instituto de Ciencias Políticas de la Universidad de Zúrich.

Contenido relacionado

Artículos similares:



Acceso restringido

[Muestreo aleatorio en el diseño de corpus: generalizabilidad entre contextos en la recopilación automatizada de eventos de protesta en varios países](#)

Mostrar detalles ▾



Acceso disponible

[Métodos innovadores en el estudio de la protesta: Introducción del editor](#)

Mostrar detalles ▾



Acceso restringido

[Análisis de organización de acciones: Ampliación del análisis de eventos de protesta mediante sitios web recuperados de centros](#)

Mostrar detalles ▾

[Ver más](#)

SAGE recomienda:

Sabio Conocimiento

Capítulo del libro

[Noticias y opiniones: diseño del entorno simbólico del mundo](#)

Mostrar detalles ▾

Sabio Conocimiento

Revisión de literatura

[Cobertura de prensa internacional de eventos de protesta en Alemania Oriental, 1989](#)

Mostrar detalles ▾

Sabio Conocimiento

Entrada

[Reuters](#)

Mostrar detalles ▾

[Ver más](#)

También de SAGE Publishing

Biblioteca CQ

Recursos políticos estadounidenses

Data Planet

Un universo de datos

SAGE Business Cases

Casos del mundo real al alcance de su mano

SAGE Campus

Cursos de habilidades y métodos en línea

SAGE Knowledge

La mejor biblioteca de ciencias sociales

SAGE Research Methods

La biblioteca de métodos definitiva

Colecciones de video SAGE Video

Streaming

Tecnología de SAGE

Facilita el aprendizaje y la investigación