

Una metodología para modelar fenómenos culturales latentes: interés público en la red *Razón Pública**

A methodology for Modeling Latent Cultural Phenomena: Public Interest in *Razón Pública*'s Network

Uma metodologia para modelar fenômenos culturais latentes: interesse público na rede *Razón Pública*

Emilio Calderón Reyes^a
Western University, Canadá
ecalder@uwo.ca

DOI: <https://doi.org/10.11144/Javeriana.uh90.mmfc>

ORCID: <https://orcid.org/0000-0003-0528-1975>

Recibido: 08 diciembre 2019

Aceptado: 03 abril 2020

Publicado: 30 diciembre 2021

Resumen:

Este artículo amplía el horizonte de análisis en el campo de las redes culturales, al proponer una metodología basada en datos para modelar fenómenos culturales latentes. A modo de ilustración, se aplicó el diseño metodológico en un estudio sobre los temas de interés público en la revista *Razón Pública*. Primero, se revisó el término interés público y se justificó la pertinencia de concebir este tipo de temas como fenómenos culturales latentes en la red de la publicación. Segundo, se presentó el periodismo digital en Colombia como el dominio cultural de la red. Para la elaboración del modelo, se propuso una estrategia para delimitar, obtener y estructurar los nodos de los actores y objetos culturales. Finalmente, se explicó el uso de la técnica Asignación Latente de Dirichlet (LDA) para modelar los temas de la revista y se validó con una encuesta que confirmó los temas de interés público, según algunos actores de la red.

Palabras clave: interés público, esfera pública, periodismo digital, red cultural, *Razón Pública*, fenómeno cultural latente.

Abstract:

This article broadens the horizon of analysis in the field of cultural networks, by proposing a data-based methodology to model latent cultural phenomena. By way of illustration, the methodological design was applied to a study on topics of public interest in the digital magazine *Razón Pública*. First, the term public interest was reviewed and justified the relevance of conceiving this type of public interest issues as latent cultural phenomena. Then digital journalism in Colombia was presented as the cultural domain of the network. For the elaboration of the model, a strategy was proposed to delimit, obtain and structure the nodes of the actors and cultural objects. Finally, it was explained how the Latent Dirichlet Allocation (LDA) technique will be used to model the magazine's topics and it will be validated with a survey if such topics are of public interest to some actors in the network.

Keywords: Public interest, public sphere, digital journalism, cultural network, latent cultural phenomenon.

Resumo:

Este artigo amplia o horizonte de análise no campo das redes culturais, ao propor uma metodologia baseada em dados para modelar fenômenos culturais latentes. A título de ilustração, o desenho metodológico foi aplicado em um estudo sobre temas de interesse público na revista *Razón Pública*. Primeiramente, revisou-se o termo interesse público e justificou-se a relevância de conceber este tipo de questões como fenômenos culturais latentes na rede de publicação. Em segundo lugar, o jornalismo digital na Colômbia foi apresentado como domínio cultural da rede. Para a elaboração do modelo, foi proposta uma estratégia para delimitar, obter e estruturar os nós dos atores e objetos culturais. Por fim, foi explicado como a técnica Latent Dirichlet Allocation (LDA) é utilizada para modelar os tópicos da revista e é validada com um levantamento se tais tópicos são de interesse público, segundo alguns atores da rede.

Palavras-chave: fenômeno de interesse público, esfera pública, jornalismo digital, rede cultural, razão pública, fenômeno cultural latente.

Notas de autor

^a Autor de correspondencia. Correo electrónico: ecalder@uwo.ca

Introducción

Razón Pública es una revista digital colombiana fundada en 2008 por un grupo de intelectuales con la misión de promover el periodismo crítico e independiente. La revista publica cada semana análisis de actualidad política escritos, mayoritariamente, por académicos e investigadores. La propuesta editorial prioriza la calidad sobre la cantidad, así como la profundidad y la reflexión sobre la inmediatez (Gómez Buendía, 2018). Los objetivos centrales son: 1) publicar análisis rigurosos y contextualizados de actualidad, 2) ser un punto de convergencia de las comunidades académicas para reflexionar sobre Colombia y el mundo, 3) divulgar los análisis entre quienes estén interesados, especialmente, entre los líderes de opinión del país (*Razón Pública*, 2009).

DiMaggio (2014) presenta las ventajas de los análisis de redes como metodologías adecuadas para hacer estudios empíricos sobre fenómenos culturales desde teorías contemporáneas relacionales, como, por ejemplo, la de los campos de Bourdieu. Según Suárez et al. (2015), como mínimo, una red cultural se compone de nodos de actores, objetos culturales y las conexiones directas entre estos. Este artículo expande las posibilidades de dicho esquema básico al incluir fenómenos culturales latentes; entendidos como un tipo de nodos que, a la vez, permean los objetos culturales de una red, pueden ser abstraídos por los actores de esta y modelados a partir de análisis de datos. La segunda sección propone una metodología basada en la definición de fenómenos culturales latentes para integrarlos en un modelo de datos de una red cultural. Las secciones posteriores ejemplifican cómo implementar dicha metodología en un estudio sobre los temas de interés público en la red de *Razón Pública*.

Metodología para formular modelos de datos con fenómenos culturales latentes

Brown et al. (2016a) establecen una secuencia de tres etapas para el análisis de redes: 1) proponer un modelo de datos basado en la información disponible y su estructura, 2) limpiar los datos con atención a eliminar duplicados y asegurar que no haya inconsistencias de formato, y 3) transformar los datos mediante programación u otros medios conforme al modelo de datos propuesto. Respecto a este proceso, la metodología propuesta en este artículo se centra en la primera etapa: formular el modelo de datos.

La delimitación de una red, como proceso asociado con fijar sus fronteras y decidir qué tipos de nodos incluir, es primordial para proponer un modelo de datos. Aunque a veces es posible reconstruir una red en su totalidad, en otros casos resulta necesario establecer criterios de delimitación. Precisamente, la definición del tipo de nodos que se contemplan es la primera decisión metodológica para delimitar una red. Normalmente los esquemas de datos tienen un solo tipo de nodos (Marin y Wellman, 2014), como es el caso de las redes sociales, conformadas exclusivamente por personas y sus relaciones directas. Con menor frecuencia, se investigan redes compuestas por dos tipos de nodos diferentes, por ejemplo, las redes de afiliaciones incluyen personas e instituciones con las que se relacionan, ya sea en términos laborales, educativos o de otra índole.

Como se mencionó, una red cultural debe tener al menos un tipo de nodos de objetos culturales y otro de actores. Dicho esquema básico se puede ampliar, al incorporar en un mismo modelo de datos múltiples tipos de objetos culturales o actores. En un estudio sobre redes culturales en El Siglo de Oro, Brown et al. (2016b) usaron un modelo de datos inicial con cuatro tipos de nodos de objetos culturales (documentos, ediciones, manuscritos y tiradas de impresión) y actores con múltiples roles (autores, impresores, mecenas y signatarios), para analizar los preliminares de libros impresos en el Imperio Español, entre 1500 y 1520.

Con el objetivo de continuar ampliando el horizonte de análisis de las redes culturales, este artículo plantea la posibilidad de incluir en modelos de datos un tipo de nodos adicional que abarque los fenómenos culturales latentes, ya definidos. La decisión de si asociar o no fenómenos culturales en el modelado de una red debe ser

evaluada teniendo en cuenta: el dominio cultural en el que esta se inscribe, los objetivos y el marco teórico de investigación; la factibilidad de extraer dichos fenómenos latentes mediante un análisis riguroso de los objetos culturales, y la posibilidad de validar con actores de la red sus perspectivas frente a esos fenómenos. A partir de estas consideraciones, se propone la siguiente secuencia de fases para la construcción de un modelo de datos de una red cultural, en torno a un tipo de fenómeno cultural latente: 1) contextualizar el dominio cultural específico en el que se inscribe la red; 2) exponer y justificar la pertinencia de estudiar el fenómeno cultural que se presume latente; 3) seleccionar y describir los tipos de objetos culturales y actores incluidos en el modelo de datos; 4) establecer pautas generales para obtener y procesar los datos con los que se modelan los objetos culturales y actores de la red; 5) proponer qué el método para modelar los fenómenos culturales latentes, a partir del análisis de los objetos culturales de la red; 6) plantear un método para validar que los fenómenos culturales latentes extraídos sean considerados como tales por una muestra de actores de la red.

Para explicar con mayor detalle en qué consiste la metodología, se ejemplifica en un estudio sobre interés público en la red de la revista *Razón Pública*. Cabe señalar que, si bien la metodología es novedosa en el campo de las redes culturales por la inclusión de fenómenos culturales latentes, las etapas están basadas en los lineamientos generales propuestos por Suárez et al. (2015).

El dominio cultural: periodismo digital en Colombia

El primer paso de la metodología consiste en identificar el dominio cultural en el que se inscribe la red, que, en el caso de *Razón Pública*, es el periodismo digital en Colombia. Para empezar, a partir de 2006 se ha consolidado el periodismo centrado en lo digital en América Latina, en términos de la cantidad de nuevos proyectos, heterogeneidad de líneas editoriales, formatos, modelos de negocio y audiencias (Meléndez Yúdice, 2015). Con relación al dinamismo que hay en el campo, este sigue siendo un fenómeno que no se ha estudiado lo suficiente (Salaverría, 2016).

Buena parte de la investigación sobre periodismo digital colombiano se ha centrado en cubrir aspectos amplios a nivel nacional, entre estos: la historia de los medios de comunicación en línea (Rey y Huertas, 2010), el contexto tecnológico y social (Gutiérrez Coba y García Perdomo, 2016), innovación y tendencias (Rey y Huertas, 2012; Zuluaga y Martínez, 2012) e interactividad (Barredo-Ibáñez y Díaz-Cerveró, 2017). Una conclusión recurrente en estos estudios es que, tras la popularización de Internet, aproximadamente en 2006, este campo ha entrado en un periodo de expansión, a pesar de las dificultades de financiación y consolidación de audiencias.

La historia de este tipo de periodismo en Colombia se remonta a la aparición de las primeras páginas web en 1996, en los siguiente diez años hubo un periodo de experimentación y lento crecimiento (Rey y Huertas, 2010). En el año 2005, la conectividad llegó al 11.01% de la población y, a partir de ese momento, la penetración de Internet se aceleró notablemente hasta llegar al 55.9% de la población en el 2015 (International Communication Union, 2019). Considerando dicha expansión, no sorprende que Rey y Huertas (2010) identificaran que, desde el 2006 se haya dado un auge de múltiples iniciativas de periodismo digital en Colombia.

Alexa Amazon ofrece una herramienta para comparar la popularidad de sitios web, con base en métricas como el número diario de páginas vistas por usuario, duración promedio de las visitas, orígenes del tráfico y el total de otras páginas que la vinculan. En Colombia, las páginas más visitadas de periodismo son: Minuto30.com (séptima posición), Pulzo.com (octava), CanalRCN.com (décima), ElTiempo.com (decimoctava), Las2Orillas.com (puesto 33), CaracolTv.com (41) y ElEspectador.com (42). En este top aparecen tanto proyectos nativos digitales, como Minuto30 y Pulzo, en los primeros lugares, y medios tradicionales que han migrado a lo digital, como Caracol y RCN. En comparación con estas páginas top, *Razón Pública* está en la posición 5.888, un resultado entendible por su hiperespecialización en análisis

político de actualidad. Mientras que páginas como Pulzo, Minuto30 y Canal RCN cubren un panorama amplio de temas y tienen audiencias masivas, *Razón Pública* se asemeja al modelo de las *single subject communities* propuesto por Nolan y Setrakian (2014), porque apela a una audiencia de nicho con contenidos que satisfagan expectativas de análisis de profundidad e independencia.

Teniendo en cuenta la misión de *Razón Pública* de influir a los líderes de opinión, los resultados del Panel de Opinión más reciente de la firma Cifras y Conceptos (2019) complementan la perspectiva presentada por Alexa. En estas páginas más mencionadas por líderes de opinión como sus fuentes de información fueron: *El Tiempo* (18%), *Semana* (16%), *La Silla Vacía* (15%), *El Espectador* (13%), *Las 2 Orillas* (6%). *Razón Pública* fue referida por el 3% de los participantes, en empate en el sexto puesto con *El País de España*, *Portafolio* y *El País de Cali*.

Temas de interés público como fenómenos culturales latentes

Esta sección desarrolla el segundo punto de la metodología propuesta. Se presenta un marco teórico sobre el interés público y se justifica la pertinencia de estudiar ese tipo de temas como fenómenos culturales latentes en la red *Razón Pública*. En este caso, se toma como base la revisión de Nancy Fraser (1990) al concepto de interés público, especialmente, en su propuesta de investigarlo inductivamente.

Habermas (1989) popularizó el concepto de la esfera pública con la traducción al inglés de *Las transformaciones de la esfera pública*. Originalmente, el autor usó dicho término en un estricto sentido histórico-sociológico para referirse al auge y declive de las organizaciones burguesas en Europa, en los siglos XVIII y XIX, como contrapartes a las autoridades estatales. Según Habermas (1989), en la esfera pública se daba un tipo de interacción distinta tanto de la vida privada como del ámbito de las autoridades estatales. Específicamente, se promovía que los participantes deliberaran libremente, como si fueran iguales, a pesar de las posibles diferencias de estatus social entre ellos.

Fraser (1990) plantea una postura crítica frente a la formulación clásica de la esfera pública de Habermas. La autora cuestionó al filósofo alemán por idealizar las instituciones burguesas y no problematizar que excluían a las minorías, particularmente a las mujeres y a las clases sociales bajas. Fraser consideró que el concepto de la esfera pública es relevante para el análisis crítico de las democracias contemporáneas, siempre y cuando, se revisen algunas de las premisas iniciales de Habermas.

Para el propósito de este artículo, se retoma esta crítica, dado que, en *Las transformaciones estructurales de la esfera pública*, Habermas no reconoció que la delimitación entre lo público y lo privado es porosa y está influenciada por las relaciones de poder entre los participantes. Como respuesta, Fraser propone entender que el interés público está constituido, alterado, disputado y negociado discursivamente en la esfera pública (p. 71). También cuestiona que se asuma que quienes participan en una misma esfera pública tienen una noción compartida sobre el interés público. Incluso, tras deliberar intensamente, es posible que no lleguen a una visión convergente o consensuada respecto a la delimitación del interés público. Justamente, la autora argumenta que, en ocasiones, la deliberación política se centra en debates sobre qué temas merecen ser considerados como de interés público, por ejemplo, algunos de los logros del feminismo radicarón, justamente, en llevar a la palestra temas que estaban circunscritos al ámbito de la vida privada, como la violencia doméstica.

A modo de precisión, las acepciones principales del término *publicidad* según Fraser son las siguientes: 1) relacionado con el Estado, 2) accesible para todos, 3) concerniente a todos, 4) relacionado con el bien común o un interés compartido. Vale la pena considerar estas acepciones en relación con el proyecto *Razón Pública*. La primera guarda una relación directa con la revista por su énfasis en el análisis político y de actualidad; por eso, se espera que las publicaciones de esta red cubran temas relacionados con el Gobierno, las diferentes ramas del poder público y la competencia por el poder político, especialmente en el contexto colombiano. En segundo lugar, la página web de *Razón Pública* es de libre acceso y cualquiera con una conexión a Internet

puede navegarla sin necesidad de registrarse. Sin embargo, para el propósito de este artículo, hay que precisar que la accesibilidad a la red cultural de *Razón Pública* no es solamente cuestión de quiénes y cómo entran a la página web, sino que también hay que considerar otros roles en los que se accede a esta como espacio deliberativo, por ejemplo, como autor, editor, promotor o facilitador. La tercera y cuarta acepciones de la publicidad son deícticas por tener como marcos de referencia implícitos a los actores que deliberan. Por un lado, los participantes interpretan la actualidad desde sus propias perspectivas y basados en estas proyectan una visión sobre lo que consideran son o deben ser las prioridades de la sociedad en su conjunto. Por lo tanto, para trazar lo público en la red cultural de *Razón Pública* no basta con estudiar los temas de la revista, sino que también es necesario caracterizar a sus actores principales.

Rauchfleisch (2017) rastreó la expansión del concepto de esfera pública en un análisis bibliométrico de 5386 artículos publicados entre 1996 y 2014. Los resultados revelan el amplio espectro de campos en los que se ha retomado el concepto, el autor concluye que la esfera pública es un concepto esencialmente disputado, pues no hay un consenso entre los múltiples autores que acuden a este sobre cómo delimitarlo y definirlo. Según este estudio, uno de los campos de investigación con mayor vigencia es el de las relaciones entre esferas públicas, tecnologías de la información y redes sociales. Es preciso profundizar en la deliberación sobre los asuntos públicos en medios digitales debido a que estos son la principal fuente de información noticiosa a nivel global (Reuters Media Institute 2018). Dahlgren (2009) concibe una esfera pública funcional como una constelación de espacios comunicativos que posibilitan la deliberación, la circulación de ideas en una sociedad y, en últimas, la formación de la opinión pública.

El término interés público proviene de la filosofía política y se ha extendido a varios campos, como el derecho, la sociología y el periodismo. Tanto Johnston (2017) como Wheeler (2006) señalan que, si bien el interés público es un término ampliamente usado, ha sido difícil de definir y delimitar. Para explorar la investigación respecto al interés público, es importante remontarse a las cuatro corrientes identificadas por Cochran (1974), que según Bozeman (2007) y Johnston (2017) siguen vigentes en las ciencias sociales: 1) abolicionista, abandonar el concepto de interés público por lo difícil que resulta definirlo y delimitarlo; 2) normativa, optar por definiciones a priori sobre el interés público; 3) comunitaria-consensualista, entender el interés público como la agregación de los intereses individuales de los miembros de una comunidad, y 4) procedimentalista, concebir el interés público como un derivado de la deliberación política.

Aunque el interés público tenga problemas de conceptualización, este es fundamental para analizar *Razón Pública* pues es un eje de su propuesta editorial (*Razón Pública*, 2009). Las críticas de Fraser (1990) a los enfoques normativos sobre el interés público son razón para optar por un enfoque inductivo y basado en datos para explorar dicho fenómeno en el caso específico de la red *Razón Pública*. El conjunto de autores, editores y directivos de la revista han contribuido de diferentes formas a la línea editorial de la revista. Sin embargo, resulta difícil rastrear los procesos deliberativos de una red digital centrada en las publicaciones de autores y en la cual la interacción de los usuarios/lectores es poco visible. Si bien los procesos deliberativos subyacen a la red *Razón Pública* y a la construcción de lo que sus miembros consideran de interés público, recopilar los datos de ese proceso es un desafío metodológico. De entre los cuatro enfoques mencionados, el comunitario-consensualista es el más apropiado respecto al propósito de realizar un estudio empírico sobre la red *Razón Pública*. Por medio del análisis de los temas prevalentes de los objetos culturales de *Razón Pública* y las perspectivas de una muestra de actores de la red, se rastrea el interés público como un fenómeno cultural latente en dicha red. Además, se tiene en cuenta quiénes acceden y participan en esta, y su grado de semejanza.

Frente a la relación entre interés público, periodismo y *Razón Pública* es pertinente referenciar *El Conflicto, callejón con salida*, de Gómez Buendía et al. (2003). El documento, que antecede la fundación de la revista, señala las falencias de los medios de comunicación colombianos respecto al cubrimiento del conflicto armado, especialmente, la falta de análisis, profundidad y contextualización. Ante dicho fenómeno, se argumenta que le corresponde al periodismo servir a los intereses públicos, fomentar la deliberación política, la conciliación y no simplificar la compleja actualidad nacional. Considerando que promover el periodismo de interés público

es un elemento central de la misión de *Razón Pública*, resulta necesario revisar los conceptos de esfera pública e interés público para dar cuenta de las especificidades de esta red. Por último, cabe señalar que el modelo toma un enfoque inductivo, en el sentido de estudiar empíricamente qué se concibe como de interés público en el contexto de una red específica.

Construcción del modelo de interés público en la red Razón Pública

Tras la revisión del dominio cultural de *Razón Pública* y la delimitación del interés público como fenómeno cultural latente que se busca abarcar, esta sección agrupa las fases tres a seis de la metodología planteada (obtención, limpieza y procesamiento de los datos) aplicada a la formulación de un modelo de datos de *Razón Pública* como red cultural.

Delimitación de tipos de objetos culturales y actores

Para analizar un fenómeno cultural latente en una red es necesario empezar por delimitar qué tipos de objetos culturales y actores se incluyen en el modelo. Debido a que la recolección, limpieza y análisis de datos siempre implica una carga de trabajo, se consideró aceptable excluir tipos de nodos que sean poco numerosos, redundantes o secundarios. Reconociendo la complejidad de dar cuenta de la totalidad de componentes y relaciones de la red *Razón Pública*, es necesario llegar a un arreglo entre la obtención de los datos y la cobertura de la multiplicidad de actores, objetos y fenómenos culturales que componen la red.

En una primera revisión de la página *Razón Pública*, el investigador navegó las secciones del menú principal y encontró los siguientes tipos de objetos culturales: 1) caricaturas en la sección *Sazón Pública*, por Anyelik, Papeto, Harpagon, Yeyonet, entre otros; 2) fotografías en Mirada pública; 3) otros materiales audiovisuales en las secciones *Videocolumnas*, *Publicast* y *Caleidoscopio*; y 4) artículos de texto. Integrar los cuatro tipos de objetos culturales en un mismo modelo complicaría el análisis, puesto que se deberían aplicar diferentes métodos para procesar cada tipo de datos, trabajo que de poco serviría, si no se lograran completar los resultados o no fueran equiparables. Por lo tanto, de entre los tipos de objetos culturales disponibles, se seleccionaron los artículos debido a que son el eje de contenido de cada número y son centrales en la misión de la revista de presentar análisis argumentados escritos por académicos.

Tras revisar los estatutos de la Fundación *Razón Pública* (2016), se estableció que los siguientes son los principales roles en los que un actor puede participar en la red: autor, lector/usuario, fundador, empleado, directivo, donante. La decisión sobre qué tipos de actores incluir en el modelo se toma de acuerdo con la posibilidad de reconstruir las relaciones entre estos y los objetos culturales ya seleccionados. Al respecto, la autoría es la única relación directa con la cual se pueden generar datos entre textos y actores. El consejo editorial de *Razón Pública* tiene una gran incidencia en lo que se publica en la revista, pues planean, editan y aprueban cada número. Sin embargo, trazar quiénes y de qué formas intervinieron en la edición de cada texto requeriría documentación sobre los procesos internos. A pesar de la complejidad de incluir los procesos editoriales en el modelo de la red, conviene señalar que las publicaciones de los autores tienen que ceñirse a las líneas y estándares del consejo editorial. Considerando que la misión de *Razón Pública* es publicar periodismo analítico sobre temas de interés público (*Razón Pública*, 2009), se puede asumir que los temas más cubiertos sintetizan una visión sobre el interés público del agregado de autores y del consejo editorial.

Por último, se propone un método para que, una vez identificados los temas latentes del conjunto de objetos culturales, se valide con actores de la red si consideran estos de interés público. Además, es necesario generar una estrategia para caracterizar la audiencia o el público alcanzados por la revista. Teniendo en cuenta que las perspectivas de los autores y el consejo editorial están representadas a nivel agregado en el corpus de artículos,

resulta pertinente que sean los usuarios/lectores de *Razón Pública* quienes validen si los temas tratados por la revista son de interés público para ellos.

Ahora bien, los lectores/usuarios son un grupo amplio que cubre desde los visitantes esporádicos hasta los seguidores más asiduos. Además, como *Razón Pública* no guarda un registro de quiénes y con qué frecuencia acceden a la página, resulta necesario acotar un segmento de usuarios regulares, que puedan ser contactados. Por lo tanto, se delimitó ese grupo a los suscriptores del boletín de la revista.

Obtención y procesamiento de datos para modelar nodos de objetos culturales, autores y suscriptores

El primer paso de la construcción del corpus de artículos de *Razón Pública* fue generar una copia espejo o *mirror* de *RazónPública.com*, actualizada a octubre 10 de 2018 con el programa *HttpTracker*. Para no sobrecargar el servidor de *Razón Pública*, el programa fue configurado con el menor número de conexiones posible, se filtraron los enlaces que no estuvieran en HTML y se estableció un límite de un 1 megabit por archivo.

La copia espejo consta de 23.758 archivos en formato HTML organizados en múltiples carpetas. Cada una fue revisada manualmente y todos los artículos fueron reunidos en una misma carpeta. Después se usó un *script* de Python (versión 3.6) con la librería *BeautifulSoup 3.0* para consolidar en un documento CSV la siguiente información por cada artículo: título, fecha, autor, palabras clave, texto completo y vistas de página. El corpus final consta de 4605 artículos tras la eliminación de elementos duplicados, incompletos o publicados de agosto del 2018, en adelante.

A partir de este corpus, se estableció una lista con los nombres de todos los autores y la cantidad de publicaciones de cada uno. La lista de nombres se revisó manualmente para anotar el género de cada autor y verificar que ningún autor estuviese duplicado. Luego, se obtuvo un consolidado con los nombres de 1152 autores de *Razón Pública* y la siguiente información de cada uno: género, total de publicaciones y títulos de los artículos de su autoría.

Con la colaboración de *Razón Pública*, fue posible distribuir entre su lista de suscriptores el enlace a una encuesta anónima cuyo objetivo era generar una caracterización. Las preguntas demográficas de la encuesta fueron: edad, género, máximo nivel de educación alcanzado, lugar de residencia (país y municipio o ciudad), actividad principal, estrato socioeconómico del lugar de residencia (únicamente para residentes en Colombia).

Método para para identificar los temas prevalentes de *Razón Pública*

La decisión de que el único tipo de objeto cultural del modelo sean artículos de texto hace factible extraer sistemáticamente los temas latentes en estos mediante la Asignación Latente de Dirichlet (LDA, por sus siglas en inglés), ideada por Blei et al. (2003). LDA es una técnica computacional para modelar los temas en corpus de textos y se ha usado para identificar cambios en líneas editoriales de medios en estudios longitudinales (DiMaggio et al., 2013) y comparar el cubrimiento y encuadre de temas específicos en diferentes medios (Ylä-Anttila et al., 2018). Esta permite construir modelos probabilísticos de un corpus que parte de la premisa de que una alta coocurrencia de términos en diferentes documentos es indicio de un tema latente. Sus creadores sintetizan la idea central del modelo de la siguiente forma: “The basic idea is that documents are represented as random mixtures over latent topics, where each topic is characterized by a distribution over words” (Blei et al., 2003, p. 993). Este tipo de modelo usa una matriz dispersa de Dirichlet para calcular la probabilidad en la que cada una de las palabras en el corpus coocurre en un mismo texto con todas las demás palabras. Los temas obtenidos tras ejecutar el algoritmo en múltiples iteraciones sirven para identificar *clusters* de palabras

que tienden a aparecer en los mismos documentos, por lo que se asume que se relacionan con un mismo tema latente.

LDA le da flexibilidad al investigador para detectar modelos de temas interpretables y coherentes en colecciones grandes de textos. Maier et al. (2018) sostienen que dicha técnica tiene un gran potencial para analizar corpus de textos cuando no es factible leer y anotar individualmente todos los documentos. Además, los autores presentan una guía de buenas prácticas para validar modelos de temas LDA, mediante la combinación de métricas cuantitativas, como la coherencia UMass y el tamaño relativo de los temas, y evaluación sustantiva, por ejemplo, examinar si las listas de palabras prevalente de cada tema se refieren a un mismo concepto o encuadre. Siguiendo a dichos autores, para generar un modelo de temas LDA a partir del corpus de artículos de *Razón Pública*, se deberían cubrir los siguientes cuatro requisitos: 1) garantizar un preprocesamiento adecuado de los datos, 2) ajustar el algoritmo con una combinación de parámetros idónea, 3) evaluar la confiabilidad e interpretabilidad del modelo, y 4) validar individualmente cada tema del modelo.

Si bien la técnica LDA requiere un proceso de validación exhaustivo, el esfuerzo se ve justificado, en este caso, para extraer de forma sistemática los temas latentes del corpus de artículos. Una vez generado y validado el modelo, es posible usarlo para establecer cuál es el tema prevalente de cada uno de los textos en el corpus. Con esos resultados, se puede trazar la relación entre cada texto y su tema prevalente y, por lo tanto, se puede relacionar el modelo de datos con cada objeto cultural y con uno de los temas latentes.

¿Cómo se valida si los temas latentes del corpus de artículos son considerados de interés público por los suscriptores?

Por último, dado que no es posible rastrear la totalidad del público alcanzado por *Razón Pública*, la encuesta para suscriptores mencionada también indagó respecto a cómo definen el interés público los suscriptores, y evaluó el nivel de consenso de esta red en torno a concepciones del interés público. En el formulario los participantes marcan el grado de interés público de los títulos de los temas previamente obtenidos en el modelo LDA entre tres opciones: 1) alto interés público, 2) interés público moderado, o 3) bajo o ningún interés público. De forma complementaria, se les pide dar tres ejemplos de temas que consideran de alto interés público. Para que los ejemplos de los participantes no estén influidos por los títulos de los temas que evalúan, se decidió que el primer elemento en el orden del cuestionario sea la pregunta abierta sobre los ejemplos de temas de interés público. Además, se les preguntó a los participantes si estaban de acuerdo o en desacuerdo con el enunciado: “En general, los medios de comunicación colombianos ofrecen un cubrimiento apropiado de los temas de interés público”. El propósito de dicha pregunta es contar con una perspectiva sobre las actitudes de los participantes respecto a la relación entre periodismo e interés público en el contexto colombiano.

Conclusiones

Este artículo presentó una metodología para desentrañar fenómenos culturales latentes y la aplicó en un estudio sobre los temas de interés público en la red de la revista *Razón Pública*. Además de la relevancia del modelo de datos propuesto para continuar un proyecto de investigación sobre una revista digital específica, la metodología puede ser útil como referencia en indagaciones futuras que impliquen modelar redes culturales. La mayor novedad del modelo es que se propone generar datos sobre un fenómeno cultural latente y desentrañar cómo está conectado con los actores y objetos culturales de una red.

El modelo de datos propuesto (Fig. 1) consta de dos tipos de nodos de actores (autores y suscriptores), uno de objetos culturales (artículos), y un cuarto nodo de fenómenos culturales latentes (temas de interés público). A grandes rasgos, la metodología consistió en contextualizar el dominio cultural de la red, presentar una aproximación teórica al concepto de interés público, examinar las especificidades de la red, establecer qué

tipos de nodos y relaciones incluir en el modelo, además de indicar cómo se obtendrían y procesarían los datos necesarios para representarlos.

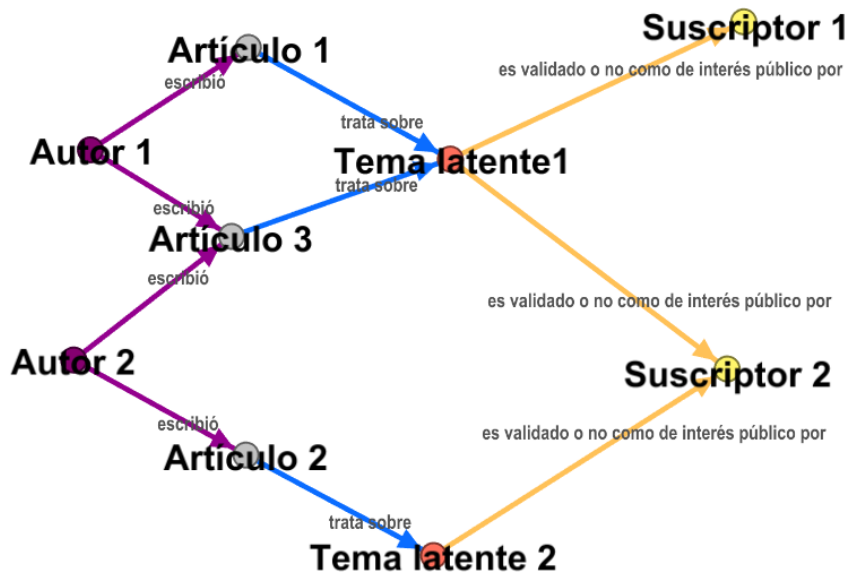


FIGURA 1
Esquema del modelo de datos de *Razón Pública*
Fuente: elaboración propia.

Proponer un modelo de datos es la primera fase para analizar una red y, de acuerdo con cómo se configure, se abren posibilidades para indagar respecto a la estructura de esta, las características y conexiones entre sus componentes. Con relación al modelo de *Razón Pública*, después de obtener y procesar los datos requeridos, se analiza la red con miras a responder las siguientes preguntas de investigación: 1) ¿cuáles son las características demográficas y profesionales, a nivel agregado, de los suscriptores y autores de *Razón Pública*? 2) ¿Cuáles son los temas latentes con mayor cubrimiento en los artículos de *Razón Pública*? 3) ¿Frente a qué temas latentes del corpus de artículos de *Razón Pública* hay mayor consenso entre los suscriptores sobre su nivel de interés público? 4) ¿Hay alguna correlación entre las características demográficas de los participantes y sus valoraciones sobre los temas latentes de los artículos?

Referencias

- Alexa (2018). *Top sites in Colombia*. <https://www.alexa.com/topsites/countries/CO>
- Barredo-Ibáñez, D. y Díaz-Cerveró, E. (2017). La interactividad en el periodismo digital latinoamericano. Un análisis de los principales cibermedios de Colombia, México y Ecuador. *Revista Latina de Comunicación Social*, (72), 273-294. <http://www.revistalatinacs.org/072paper/1165/15es.html>
- Blei, D. M., Ng, A. Y., y Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, (3), 993-1022. <http://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf>

- Bozeman, B. (2007). *Public values and public interest: Counterbalancing economic individualism*. Georgetown University Press.
- Brown, D., Soto-Corominas, A., Suárez, J. L., y De la Rosa, J. (2016a). Chapter 9: Overview – The Social Media Data Processing Pipeline. In L. Sloan, *The SAGE Handbook of Social Media Research Methods*. <http://dx.doi.org/10.4135/9781473983847.n9>
- Brown, D., Soto-Corominas, A., y Suárez, J. (2016b). The Preliminaries Project: Geography, Networks, and Publication in the Spanish Golden Age. *Digital Scholarship in the Humanities*, 32(4), 709–732. <https://doi.org/10.1093/lc/fqw036>
- Cifras y Conceptos (2019). *Panel de opinión*. Bogotá. <http://cifrasycconceptos.com/productos-panel-de-opinion/>
- Cochran, C. (1974). Political Science and “The Public Interest”. *The Journal of Politics*, 36(2), 327-355. www.jstor.org/stable/2129473
- Dahlgren, P. (2009). *Media and Political Engagement: Citizens, Communications and Democracy*. Cambridge University Press.
- DiMaggio, P. (2014). Cultural Networks. In J. Scott y P. J. Carrington, *The SAGE handbook of social network analysis* (pp. 286-300). SAGE Publications. DOI: 10.4135/9781446294413.n20
- DiMaggio, P., Nag, M., y Blei, D. M. (2013). Exploiting Affinities between Topic Modeling and The Sociological Perspective on Culture: Application to Newspaper Coverage Of U.S. Government Arts Funding. *Poetics*, (41), 570-606. <https://doi.org/10.1016/j.poetic.2013.08.004>
- Fraser, N. (1990). Rethinking the Public Sphere: A Contribution to the Critique of Actually Existing Democracy. *Social Text*, (25/26), 56-80. https://carbonfarm.us/amap/fraser_public.pdf
- Gómez Buendía, H. (2018). La inteligencia de Colombia nos cree. *El Espectador* <https://www.elespectador.com/entrenamiento/un-chat-con/hernando-gomez-buendia-la-inteligencia-de-colombia-nos-cree-articulo-819206>
- Gómez Buendía, H., Vicente, C. y Franche, M. A. (2003). El conflicto, callejón con salida. <http://hdl.handle.net/20.500.12010/13541>
- Gutiérrez Coba, L. y García Perdomo, V. (2016). Colombia. En R. Salaverría, *Ciberperiodismo en Iberoamérica* (pp. 79-96). Fundación Telefónica, Editorial Ariel.
- Habermas, J. (1989). *The Structural Transformation of the Public*. Cambridge MIT Press.
- International Communication Union (2019). *Percentage of individuals using the Internet*. https://www.itu.int/en/ITU-UD/Statistics/Documents/statistics/2019/Individuals_Internet_2000-2018_Jun2019.xls
- Johnston, J. (2017). The public interest: A new way of thinking for public relations? *Public Relations Inquiry*, 6(1), 5-22. <https://doi.org/10.1177/2046147X16644006>
- Maier, D., Waldherr, A., Miltner, P., Wiedemann, G., Niekler, A., Keinert, A., Pfetsch, B., Heyer, G., Reber, U., Häussler, T., Schmid-Petri, H., y Adam, S. (2018). Applying LDA Topic Modeling in Communication Research: Toward a Valid and Reliable Methodology. *Communication Methods and Measures*, 12(2-3), 93-118. <https://doi.org/10.1080/19312458.2018.1430754>
- Marin, A., y Wellman, B. (2014). Social Network Analysis: An Introduction. In J. Scott y P. J. Carrington, *The SAGE Handbook of Social Network Analysis* (pp. 11-25). SAGE Publications. DOI: 10.4135/9781446294413.n2
- Meléndez Yúdice, J. (2015). *Primer estudio de medios digitales y periodismo en américa latina*. <http://www.nodal.am/wp-content/uploads/2016/01/Primer-estudio-de-medios-digitales-y-periodismo-en-Am%C3%A9rica-Latina.pdf>
- Nolan, K., y Setrakian, L. (2014). “Seeking the single subject news model”. *Columbia Journalism School*. <http://towcenter.org/research/seeking-the-single-subject-news-model/>
- Rauchfleisch, A. (2017). The Public Sphere as an Essentially Contested Concept: A Co-Citation Analysis of The Last 20 Years of Public Sphere Research. *Communication and the Public*, 2(1), 3-18. <https://doi.org/10.1177/2057047317691054>
- Razón Pública (2009). ¿Qué es Razón Pública? <https://www.razonpublica.com/index.php/qus-razca-mainmenu-61.html>

- Razón Pública* (2016). Acta No 17 de Asamblea General Extraordinaria (reforma de los estatutos). <https://consultaimagenesob.ccb.org.co/OnBaseDMZ/docpop/docpop.aspx>
- Reuters Media Institute (2018). *Global News Report*. <http://media.digitalnewsreport.org/wp-content/uploads/2018/06/digital-news-report-2018.pdf>
- Rey, G. y Huertas, C. E. (2010) *Periodismo digital en Colombia: el quién y el cómo de los nuevos medios*. Universidad Javeriana. http://consejoderedaccion.org/webs/documentos/Estudio_medios_digitales_2010.pdf
- Rey, G. y Huertas, C. E. (2012) *Medios digitales en Colombia: agendas periodísticas y sostenibilidad*. Universidad Javeriana. http://www.consejoderedaccion.org/images/documentos/emprendimiento/estudio_medios_digitales_2012.pdf
- Salaverría, R. (2016). *Ciberperiodismo en Iberoamérica marco general*. Fundación Telefónica, Editorial Ariel.
- Suárez, J. L., McArthur, B., y Soto-Corominas, A. (2015). Cultural Networks on Culture and Computing. *International Conference on Culture and Computing, Kyoto*. <http://www.cultureplex.ca/wp-content/uploads/2017/09/074332381.pdf>
- Wheeler, C. (2006). The Public Interest: We Know it is Important, But Do We Know What It Means. *ALAL Forum*, (48), 12-25. <https://search.informit.org/doi/10.3316/IELAPA.200605460>
- Ylä-Anttila, T., Eranti, V., y Kukkonen, A. (2018). *Topic Modeling as a Method for Frame Analysis: Data Mining the Climate Change Debate in India and the USA*. <https://doi.org/10.31235/osf.io/dgc38>
- Zuluaga, J., y Martínez, M. P. (2012). *Mapping Digital Media. Colombia*. <https://www.opensocietyfoundations.org/publications/mapping-digital-media-colombia>

Notas

- * Artículo de investigación en el marco del proyecto “Interés Público en la revista *Razón Pública*”.

Licencia Creative Commons CC BY 4.0

Cómo citar este artículo: Calderón Reyes, E. (2021). Una metodología para modelar fenómenos culturales latentes: interés público en la red *Razón Pública*. *Universitas Humanística*, 90. <https://doi.org/10.11144/Javeriana.uh90.mmfc>