

ISSN 0120-4157

Biomédica

Revista del Instituto Nacional de Salud

PUBLICACIÓN ANTICIPADA EN LINEA

El Comité Editorial de *Biomédica* ya aprobó para publicación este manuscrito, teniendo en cuenta los conceptos de los pares académicos que lo evaluaron. Se publica anticipadamente en versión pdf en forma provisional con base en la última versión electrónica del manuscrito pero sin que aún haya sido diagramado ni se le haya hecho la corrección de estilo.

Siéntase libre de descargar, usar, distribuir y citar esta versión preliminar tal y como lo indicamos pero, por favor, recuerde que la versión impresa final y en formato pdf pueden ser diferentes.

Citación provisional:

Rivera-Romano LS, Juárez-Cano G, Hernández-Lemus E, Vallejo M, Martínez-

García M. Estructura de comunidades en redes semánticas de investigación

biomédica sobre disparidades en salud y sexismo. *Biomédica*. 2020;40(4).

Recibido: 22-08-19

Aceptado: 03-07-20

Publicación en línea: 07-07-20

Estructura de comunidades en redes semánticas de investigación biomédica sobre disparidades en salud y sexismo

Structure of communities in semantic networks of biomedical research on disparities in health and sexism

Comunidades en redes semánticas en investigación biomédica

Lucero Soledad Rivera-Romano ¹, Gabriela Juárez-Cano ¹, Enrique Hernández-Lemus ², Maite Vallejo ³, Mireya Martínez-García ³

¹ Programa de Licenciatura en Promoción de la Salud, Universidad Autónoma de la Ciudad de México, Ciudad de México, México

² División de Genómica Computacional, Instituto Nacional de Medicina Genómica, Ciudad de México, México

³ Departamento de Investigación Sociomédica, Instituto Nacional de Cardiología Ignacio Chávez, Ciudad de México, México

Correspondencia:

Mireya Martínez-García, Instituto Nacional de Cardiología Ignacio Chávez, Juan Badiano # 1, Sección XVI, Tlalpan, CP 14080, Ciudad de México, México.

Tel.: +52 55 73 52 55 ext. 24302.

mireya.martinez@cardiologia.org.mx

Contribución de los autores:

Lucero Soledad Rivera-Romano: diseño del estudio.

Enrique Hernández-Lemus, Maite Vallejo y Mireya Martínez-García: diseño y planificación del estudio y redacción del trabajo.

Todos los autores participaron en el análisis e interpretación de datos.

Introducción. Como una iniciativa para mejorar la calidad de la atención sanitaria se ha incrementado la tendencia en investigación biomédica centrada en las disparidades en salud y el sexo.

Objetivo. Realizar una caracterización de la evidencia científica sobre la disparidad en salud definida como la brecha existente entre la distribución de la salud y el posible sesgo por sexo en el acceso a los servicios médicos.

Materiales y métodos. Se realizó una búsqueda simultánea de la literatura científica en la base de datos MEDLINE PubMed de dos descriptores fundamentales: *Healthcare disparities* y *Sexism*. Posteriormente se construyó una red semántica principal y se identificaron algunas subunidades estructurales (comunidades) para el análisis de patrones de organización de la información. Se utilizó software de código abierto: Cytoscape para análisis y visualización de redes y MapEquation para la detección de comunidades y código desarrollado ex profeso y disponible en un repositorio de acceso público.

Resultados. El corpus de la red principal mostró que los términos sobre las enfermedades del corazón fueron los más coocurrentes entre los descriptores de condiciones médicas. De las subunidades estructurales se identificaron patrones de información relacionada con las políticas públicas, los servicios de salud, determinantes sociales y los factores de riesgo, pero con cierta tendencia a mantenerse indirectamente conectados a los nodos de condiciones médicas.

Conclusiones. La evidencia científica indica que la disparidad por sexo sí importa para la calidad de la atención de muchas enfermedades, especialmente aquellas relacionadas con el sistema circulatorio. Sin embargo, aún se percibe un

distanciamiento entre los factores médicos y sociales que dan lugar a las posibles disparidades por sexo

Palabras clave: investigación biomédica; calidad de la atención de salud; disparidades en el estado de salud; sexismo; minería de datos; interpretación estadística de datos; web semántica.

Introduction: As an initiative to improve the quality of health care, the trend in biomedical research focused on health disparities and sex has increased.

Objective: Carry out a characterization of the scientific evidence on health disparity, defined as the gap between the distribution of health and the possible gender bias in access to medical services.

Materials and methods: A simultaneous search of the scientific literature was performed in the MEDLINE PubMed database of two fundamental descriptors: Healthcare disparities and Sexism. Subsequently, a main semantic network was built and some structural subunits (communities) were identified for the analysis of information organization patterns. Open source software was used: Cytoscape to analyze and visualize the semantic network and MapEquation for community detection, as well as ad hoc code available in a public access repository.

Results: The core network corpus showed that the terms on heart disease were the most common among the descriptors of medical conditions. Patterns of information related to public policies, health services, social determinants and risk factors were identified from the structural subunits, but with a certain tendency to remain indirectly connected to the nodes of medical conditions.

Conclusions: Scientific evidence indicates that gender disparity does matter for the quality of care for many diseases, especially those related to the circulatory system. However, there is still a gap between the medical and social factors that give rise to possible disparities by sex.

Key words: Biomedical research; quality of health care; health status disparities; sexism; data mining; data interpretation; statistical; semantic web.

Reducir las brechas en la base de conocimiento científico relacionadas con las disparidades por sexo y lograr la equidad en salud debe ser una prioridad para mejorar la calidad de la atención médica. Recientemente, en la literatura biomédica se ha incrementado la comprensión de que el sexo sí importa para la forma en la que se desarrolla el proceso salud-enfermedad y la calidad de la atención médica, hecho que puede verse afectado tanto por su biología como por el contexto social (1). La política y los programas de salud global a menudo son ciegos a las diferencias entre sexos, y sobre todo, a la posición desigual de las mujeres en la sociedad, lo que hace que éstas sean *visiblemente invisibles* (2). Para comenzar, es pertinente definir de manera conceptual el sexo y el género, en seguida, también abordaremos el marco conceptual de las disparidades en salud. Por sexo se entienden las influencias biológicas, es decir, las diferencias reproductivas, hormonales y genéticas entre mujeres y hombres. El *género*, más bien, se refiere a los roles, comportamientos, actividades y atributos socialmente construidos que una sociedad considera apropiados para hombres y mujeres (3). Para fines de esta investigación, se utilizará el concepto de sexo para el análisis semántico de la información biomédica y el concepto de disparidades en salud como lo define la Biblioteca Nacional de Medicina de los Estados Unidos (MEDLINE) PubMed: Diferencias en el acceso o disponibilidad de servicios médicos.

El sexo es una característica biológica determinante para la salud, documentar ese efecto es un punto de partida obligado para la investigación de disparidades y sus hallazgos pueden facilitar el desarrollo de intervenciones diseñadas para reducirlas (4). La inclusión sexual es una cuestión de integridad científica. La

brecha en la representación de las mujeres en los estudios de ciencia básica ha sido bien documentada y hasta hace poco no se había hecho mucho por mitigarla (5-8).

Este hecho ha sido acertadamente abordado en las directrices del reporte SAGER (The Sex and Gender Equity in Research, siglas en inglés: SAGER), una guía para reportar los resultados de los estudios biomédicos por sexo y género, incorporando esta dimensión desde la planificación, el diseño del estudio, el análisis de los datos y la interpretación de los resultados a fin de promover la equidad, la ética, la responsabilidad social y el mayor aprovechamiento de la evidencia científica (9). Así mismo, la Organización Mundial de la Salud (OMS) identificó brechas de conocimiento sustanciales sobre la interacción del sexo con otros determinantes sociales de la salud, y ha recomendado que los avances en investigación biomédica no se desarticulen de esos determinantes (5).

Por otro lado, el marco conceptual en el que se encuadra la disparidad de salud mediada por el sexo se relaciona estrechamente con los determinantes sociales, ya sea en su configuración estructural (políticas intersectoriales, gobernanza socioeconómica) o mediadora (empoderamiento e inclusión a los sistemas establecidos para mantener la salud), o ambas, cuyos efectos muchas veces inciden en la distribución desigual de recursos, generando estratificación social y divisiones de clase que contribuyen a modelar las principales inequidades sociales afectando el estado de salud de hombres y mujeres (10,11).

Si bien, en las últimas décadas muchos países lograron sensibilizarse y mejorar las condiciones de salud en general, lamentablemente los beneficios de estas acciones no se han apreciado en algunos grupos poblacionales, como ha sido el

caso entre hombres y mujeres. ¿Por qué? Una suposición es que, a pesar de haber muchos estudios científicos, los datos no son reportados de forma desagregada por sexo y eso desorienta el efecto en los resultados, por ejemplo, en cuanto al reporte de determinantes sociales o disparidades en salud, o al tipo y grado de enfermedad y/o muerte (12-14).

Otra hipótesis es una discrepancia teórico-práctica en la salud pública, que no ha logrado vinculado por completo los conocimientos del paradigma biomédico (muchas veces individualista) y los determinantes sociales subyacentes al proceso salud-enfermedad, que influyen en las elecciones, los conocimientos y comportamientos de las personas y por lo tanto, en sus resultados de salud (11,15).

La medicina basada en el sexo es una rama innovadora de la investigación biomédica y representa una nueva perspectiva para la promoción de la salud y la calidad de la atención médica (16). Sin embargo, con el continuo aumento en la cantidad de literatura, es difícil para los tomadores de decisiones, médicos e investigadores utilizar los recursos disponibles y mantenerse al día de manera rápida y efectiva en este tema (17).

Aunque los sistemas electrónicos de recuperación de información proporcionan documentos potencialmente útiles, no suelen ayudar a administrar y analizar el gran volumen de información recuperada (18). Recientemente se han propuesto diversas metodologías automatizadas para extraer la información de MEDLINE PubMed y visualizarla como un grafo utilizando los términos del vocabulario controlado de Medical Subject Headings (MeSH); estos términos son representados como nodos y la información compartida entre ellos se representa

como enlaces, un método novedoso de abstracción semántica para resumir múltiples textos biomédicos (19,20).

El modelado conexionista, se ha empleado justamente para explorar las representaciones léxicas mediante las redes semánticas (21). Estas redes permiten la extracción de ideas significativas mediante la identificación de grupos de conceptos emergentes y/o desconocidos de manera inductiva (22). El análisis de redes revela propiedades de agrupamiento que proporcionan información sobre la estructura del léxico y la semántica, ninguno de los cuales es directamente observable a partir de los datos aislados (23).

Por lo tanto, el problema de investigación que nos ocupa en este estudio está basado en la conceptualización de la disparidad en salud definida como la brecha existente entre la distribución de la salud y el posible sesgo por sexo de la evidencia científica reportada en dos sentidos: 1) Aquel relacionado con los mecanismos sociales y el acceso oportuno a los sistemas de salud -- muchos de ellos evitables e injustos; 2) El relacionado con los mecanismos biológicos y el manejo médico-clínico, algunos de ellos quizá inevitables pero igualmente vinculados a las diferentes necesidades específicas de hombres y mujeres.

La pregunta a responder en esta investigación mediante el uso de redes semánticas es si la información reportada en un corpus de literatura biomédica aborda el tema de la disparidad desagregada por sexo, y si los MeSH sobre temas biomédicos se encuentran vinculados con MeSH sobre determinantes sociales relacionados con el acceso y la atención de los servicios de salud.

Por lo tanto, a partir de la minería computacional de la literatura biomédica y la teoría de grafos, el objetivo del presente estudio es realizar una caracterización de

la evidencia científica más reciente sobre las disparidades en salud y el sexismo, para posteriormente estudiar la estructura de 'comunidades' (o módulos) dentro de las redes semánticas y generar evidencia sobre la representación altamente expresiva de los principales descriptores relacionados con los patrones de información en el campo de la investigación biomédica.

Materiales y métodos

La figura 1 resume los pasos seguidos para la conducción de esta investigación, para ello se aplicaron dos estrategias de consulta bibliográfica. Por un lado, para el análisis de redes basado en términos MeSH, el 24 de octubre de 2018 se realizó minería computarizada de la literatura científica de la base de datos de MEDLINE PubMed, para ello se emplearon simultáneamente dos descriptores sin ninguna restricción: *Healthcare disparities* y *Sexism*. Por otro lado, se consultaron otros documentos de apoyo para contextualizar el problema planteado en este trabajo y reforzar la discusión.

Si bien PubMed es la mayor base de datos de literatura biomédica del mundo, tiene algunas limitaciones. Por ejemplo, cuando se le solicita la recuperación de las citas consultadas para archivarlas, el registro devuelto (por ejemplo en el formato MEDLINE para trabajar con softwares de generación de bibliografías) no incluye todos los autores que se mencionan en los documentos indexados, asimismo se omite su afiliación institucional y el país. Por otro lado, aunque PubMed contiene más de 15 millones de referencias bibliográficas provenientes de más de 4 mil revistas internacionales, omite aquellas que no se encuentran indexadas en esta base y que podrían contener evidencia científica relevante para las ciencias de la salud.

Minería computarizada de términos MeSH

Los términos MeSH fueron diseñados por un panel de expertos internacionales en un proyecto dirigido por la Biblioteca Nacional de Medicina de los Estados Unidos de América como una **Ontología**. Este hecho tiene importantes consecuencias para su uso como sistemas de clasificación hemerobibliográfica y tiene también consecuencias para el procesamiento del minado de textos asistido por computadora. Una ontología es, como se sabe, un sistema conceptual informal representado a través de una teoría lógica, es decir es un vocabulario controlado (finito) utilizado por una teoría lógica.

La ontología MeSH contiene términos semánticamente relacionados, estructurados jerárquicamente para facilitar la consulta y el análisis computacional de las publicaciones incluidas (24). Cuenta con más de 57 mil descriptores (títulos principales) y más de 80 calificadores (subtítulos temáticos), cada publicación indexada puede contener uno o varios MeSH que son asignados según el contexto de todo el documento (25).

Esto resulta relevante pues, al generarse un vocabulario controlado se evitan los sesgos de lenguaje debidos a, por ejemplo la sinonimia o, más complejo aún a la terminología contexto-dependiente. Así, por ejemplo para referirse a un crecimiento anormal de tejido debido al cáncer, algunos autores utilizarían la palabra clave “tumor”, otros “neoplasia”, “cáncer”, “nódulo maligno” y así por el estilo. En todos estos casos la clasificación ontológica descrita utilizaría el MeSH “Neoplasms” como clasificador, evitando así las ambigüedades.

Aunque las técnicas modernas de procesamiento computacional de lenguaje natural permiten la generación de diccionarios para desambiguar situaciones

similares, éstos son aún muy proclives a fallar, particularmente en léxicos altamente técnicos (como puede ser el caso aquí) y de esta manera introducir sesgos.

Por tanto, la concurrencia de términos MeSH es un recurso muy apreciado para la identificación y la recuperación rápida, precisa y confiable de información (24,25).

Por estas razones, se utilizaron los MeSH como referencia bibliométrica para la construcción de un mapa de conectividad y posteriormente la construcción de redes semánticas, donde al menos dos MeSH conectados mediante uno o más registros PMID fueran registrados.

Estrategia de búsqueda de términos MeSH

Para cumplir con nuestro objetivo de investigación, decidimos combinar en la búsqueda de MEDLINE PubMed dos términos específicos: *Healthcare disparities* y *Sexism*. Si bien en una búsqueda individual de un solo término MeSH los resultados son más numerosos, también es cierto que ese extenso número de documentos pueden estar relacionados con temáticas menos específicas. Por otro lado, a veces incluir otros criterios de búsqueda, además de los MeSH, como Title/abstract o keywords puede elevar el número de publicaciones. Sin embargo, en la mayoría de las ocasiones este método ya no es necesario, pues la búsqueda combinada de ciertos MeSH es capaz de reunir la evidencia necesaria y cumplir con el objetivo de la búsqueda, como sucedió para este estudio.

Estrategia de selección de los documentos recuperados

Los criterios de selección que se aplicaron a los documentos recuperados fueron los siguientes:

Los criterios de inclusión:

1. Que cada registro bibliográfico contenga al menos un término MeSH para establecer la conexión en la red entre cada documento.
2. El registro bibliográfico puede ser de cualquier año de publicación.
3. El registro bibliográfico puede ser de cualquier país de adscripción.
4. El registro bibliográfico puede ser de cualquier tipo de literatura científica.
5. El término MeSH puede estar en cualquier idioma.
6. El término MeSH puede estar indexado en cualquier nivel de la estructura jerárquica de anotación.
7. El término MeSH puede estar incluido en cualquiera de las categorías biomédicas de MEDLINE PubMed.

Los criterios de exclusión que se aplicaron fueron:

1. El registro bibliográfico no contiene título o resumen.
2. El contenido del registro bibliográfico no coincide con la pertinencia del problema en estudio.

Estrategia de curado de información

Se llevo a cabo un curado o escrutinio manual de los términos MeSH recuperados que serían conectados en la red con el objetivo de identificar y posiblemente eliminar aquellos términos altamente redundantes y que por sí mismos no aportaran información relevante a la búsqueda. Esto solo sucedió con el MeSH *Humans*, (común a todos los registros de investigación en salud en humanos y por tanto redundante) el cual fue de hecho eliminado antes de analizar la estructura y la conectividad de la red. Posteriormente el corpus de información fue procesado, de texto simple a lenguaje natural utilizando un código en el lenguaje de

programación Python previamente desarrollado para la construcción de una red semántica (5).

Construcción de redes semánticas

Las redes semánticas son representaciones gráficas de conocimiento basadas en relaciones significativas de ideas o conceptos, estructuradas como una red de palabras relacionadas cognitivamente entre sí. Se utilizan comúnmente para inferir marcos teóricos o conceptuales utilizados en los textos, y muchas veces su análisis contribuye a que surjan nuevos enfoques de los datos explorados (18).

La construcción de redes semánticas está basada en la teoría conexionista de redes complejas. Una red es un conjunto de nodos, elementos o vértices que se conectan mediante de aristas o enlaces que pueden ser dirigidos (que tienen una dirección de un nodo a otro) o no dirigidos. El objetivo del estudio de redes es encontrar propiedades estadísticas o topológicas las cuales caracterizan la estructura de la misma para entender las propiedades que representa. Algunos atributos topológicos de una red o medidas de centralidad utilizados en el presente trabajo se describen brevemente a continuación:

Centralidad de valor propio: es una medida de la influencia de un nodo en una red basada en el hecho de que nodos conectados de manera preferencial a nodos muy conectados tendrán una mayor influencia en la red.

Distribución de grado: es la característica más simple que puede observarse en una red, la conectividad entre nodos adyacentes.

Coefficiente de agrupamiento: estima la tendencia de agrupamiento de nodos en una red, es la probabilidad de que dos nodos puedan estar conectados por un tercer nodo y exista mayor cohesión.

Promedio de caminos más cortos: indica la longitud promedio del camino más corto entre dos nodos.

Centralidad de intermediación: es el número de caminos cortos entre dos nodos, que pasan a través de otro nodo.

Si bien existen diferentes atributos que se pueden estudiar para la determinación de los patrones de conectividad de una red, en esta investigación elegimos los parámetros más simples que se utilizan en la teoría de redes para estimar la conexión de los conceptos, este es el caso de la *centralidad de grado* (CG), el número total de enlaces de un nodo o la suma de las frecuencias o pesos de las interacciones (26). Los nodos con alta CG tienen muchos vecinos, y generalmente se encuentran con mayor frecuencia en la red que los nodos con un pequeño grado (27,28).

Con el fin de discutir similitudes y diferencias con otras medidas de centralidad relevantes, hemos incluido una tabla con la distribución de diversas centralidades para los MeSH correspondientes a padecimientos por resultar éstas de interés especial para nuestra discusión posterior. La tabla correspondiente a la lista completa de MeSH se ha incluido como Material Suplementario (Cuadro suplementario 1).

Para una mejor visualización de algunos MeSH relacionados con las principales enfermedades encontradas en el corpus de la red, se marcaron con diferentes colores y tamaños los nodos según el número de conexiones con sus principales vecinos. Para analizar y visualizar la red se utilizó el software de código abierto Cytoscape (29).

Red principal

Una vez que se realizó la minería computarizada de la literatura, se construyó una red basada en taxonomías predeterminadas –los términos MeSH-, no dirigida ni ponderada. El nodo es la unidad de análisis de la red representado en un mapa de conectividad. Los mapas de conectividad se construyeron de modo que los nodos de origen y destino sean los MeSH (que representan a los conceptos bajo análisis) que se identificaron en el corpus de los artículos científicos.

Posteriormente se trazó un enlace (arista) entre estos nodos, siempre y cuando al menos dos artículos identificados por un número único (PubMed Identifier, PMID) compartieran ambos MeSH. Los enlaces utilizados en la construcción de la red fueron los PMID de cada publicación. Este enfoque da lugar a una red que no considera el peso relativo de las conexiones. Como se discutirá más adelante dada la distribución de pesos en nuestras redes (prácticamente idéntica para cada enlace en nuestra red con pocas excepciones, si desea consultar dichas distribuciones véanse las figuras suplementarias respectivas) este hecho no afecta de manera significativa los resultados o conclusiones del presente estudio.

Subred de enfermedad cardiovascular

Como un ejemplo para explorar los patrones de conexión de información en una subred, se dividió la red principal usando la ontología MeSH de los términos relacionados con la enfermedad cardiovascular (ECV). Coloreamos de rojo aquellos nodos y enlaces asociados con ECV, en amarillo algunos términos sugerentes de determinantes sociales, en azul oscuro los relacionados con los servicios de salud, en verde aquellos que evocan las condiciones biológicas y en color azul claro los términos alusivos a otros factores de estilos y modos de vida.

Red emergente de la diferencia entre mujeres y hombres

Para inferir patrones de conexión de información por sexo, también se construyeron subredes de hombres y mujeres. Para asegurar la comparabilidad entre las subredes, éstas se filtraron en Cytoscape seleccionando el MeSH relativo al sexo (ya sea *female* o *male*) con todos sus primeros vecinos. A partir de estas subredes se realizó la diferencia de sus nodos, para conocer los términos incluidos en una y no en la otra. A la red resultado de esta diferencia se le llamó *red emergente*.

Modularidad

La modularidad es una propiedad que permite simplificar y resaltar propiedades estructurales de una red que conduce a la aparición de subunidades fuertemente interconectadas, conocidas como módulos o comunidades; esto es, son aquellas subunidades que contienen un número mayor de enlaces entre nodos dentro de la comunidad, que el número de enlaces a nodos fuera de ella (30,31).

Si bien existen diversas técnicas de detección de comunidades, no hay un solo método de optimización de la partición modular que funcione en todos los casos (32). Para esta investigación se utilizó un algoritmo de agrupación de comunidades (basado en caminatas aleatorias y teoría de la información) centrado en la interdependencia de los enlaces llamado Infomap y para la visualización modular de estas subunidades se utilizó la aplicación en línea llamada MapEquation (33,34).

Infomap es un método que se basa en la dinámica del flujo de información sobre la red más que únicamente en su estructura topológica simple. La llamada

MapEquation captura el flujo de información en la red, por lo que las

comunidades consisten en grupos de nodos entre los cuales el flujo de información persiste. A grandes rasgos, para describir el flujo de información en la red, se usa un caminata aleatoria para reportar dicho flujo (como aproximación). Por ergodicidad, asintóticamente se llega a una distribución estacionaria tanto para cuando se visitan estructuras de la red como para cuando se abandonan (dichas estructuras pueden ser nodos o módulos). Dichas distribuciones estacionarias se pueden codificar de forma optima usando la llamada codificación de Huffman en una cadena binaria. Así entonces la caminata se puede describir en términos de la teoría de la información a partir de la longitud de la cadena binaria asociada a dicha caminata (31,32,35).

Si se tuviera una partición de la red en M en módulos, se podría tener una descripción de una caminata aleatoria entre los módulos (o estructuras más relevantes) en la red. Esta caminata se puede codificar en una cadena binaria cuya longitud $L(M)$ sería mínima en el caso de tener la partición óptima de la red. Lo anterior se puede lograr mediante la ecuación del mapa (MapEquation). Así, la MapEquation aprovecha la dualidad entre encontrar la mejor estructura modular en una red y minimizar la longitud de la descripción de los movimientos de un caminante aleatorio en dicha red. Es decir, para cada partición modular dada de la red, existe un costo de información asociado a describir los movimientos del caminante aleatorio, o del flujo. Algunas particiones generan longitudes de descripción más cortas y algunas otras más largas. La partición con la longitud de descripción más corta es la que mejor captura la estructura modular de la red con respecto a la dinámica de flujo de información en la misma (33-35).

En vista de que nos interesa encontrar módulos con distancia semántica mínima,

es decir, los correspondientes a términos más estrechamente relacionados de manera local y por tratarse de redes con miles de aristas, la combinación de un algoritmo con minimización de la distancia de código y baja complejidad computacional hace del algoritmo Infomap una elección apropiada.

Finalmente, los módulos se etiquetaron con el nombre del nodo con el mayor índice PageRank (IPR), el cual de hecho es un tipo de centralidad de valor propio (35). Éste es una medida de centralidad basada en diversos algoritmos utilizados para asignar de forma numérica la relevancia de los documentos indexados por un motor de búsqueda, de esta manera, las comunidades constituyen unidades semánticas estructuradas de mayor complejidad que los términos MeSH individuales.

Resultados

Minería computarizada de la literatura

De la minería computarizada se obtuvieron 41 artículos científicos publicados entre 2011 y 2018. Los principales países que se encontraron registrados por la adscripción de al menos un artículo son: Estados Unidos, Europa, China, Brasil, México, Canadá y Sudáfrica.

Red principal

Las principales características topológicas de las redes analizadas se muestran en el cuadro 1. La estructura revela una red principal, con 301 nodos (o términos) y 3365 enlaces (o relaciones). En ella se pudo identificar una aparente tendencia a priorizar la investigación relacionada con las disparidades de sexo y algunas enfermedades, específicamente aquellas que involucran la enfermedad cardiovascular, como se aprecia en los 16 nodos de color rojo de la figura 2.

En color rosa se ven los nodos relacionados con tumores o cáncer de mama o colorrectales (4 nodos y 108 CG acumulado). En azul los relacionados con el dolor y las propiedades de analgesia o narcosis, así como los relacionados con la salud y la enfermedad mental (6 nodos y 68 CG acumulado). En verde, los relacionados con las infecciones por virus de inmunodeficiencia humana (3 nodos y 46 CG acumulado); y en menor tamaño, se aprecian los nodos en color naranja, amarillo, y turquesa relacionados con otros padecimientos o métodos médicos, quirúrgicos y ortopédicos, hernias, desnutrición, prolapso uterino y cálculos uretrales.

En la red principal también se pueden identificar algunos términos cercanos o primeros vecinos asociados a los nodos de los padecimientos y relacionados con la calidad de la atención. Por ejemplo, cerca de los nodos en color rojo (ECV) se encuentran términos no esperados relacionados con la relación médico-paciente, la toma de decisiones médicas y el manejo de la enfermedad; por otro lado, también se encontraron términos más comúnmente asociados con la supervivencia, los factores de riesgo y los estilos de vida. Cerca de los de color rosa, se pueden apreciar términos aparentemente poco vinculados con los tumores o el cáncer como estigma social, características culturales, y accesibilidad a los servicios de salud. Y cerca de los azules, sobre dolor y analgesia, se distinguen términos como la salud de la mujer y los servicios organizados para brindar atención de salud a mujeres, ofensas sexuales y política de salud, entre otros.

Los cuadros 2 y 3 muestran los resultados comparativos entre los valores de las centralidades analizadas de la red principal cuando consideramos los MeSH que corresponden a condiciones médicas patológicas. El cuadro 2 muestra que, por

ejemplo, aunque hay diferencias en las distribuciones de rango al usar diferentes medidas de centralidad (centralidad de grado, centralidad de intermediación, centralidad de cercanía) la tendencia reportada con respecto a los padecimientos cardiovasculares se mantiene. El cuadro 3, muestra una excepción notable e interesante de mencionar por sus posibles connotaciones; es el caso de VIH, cuyo rango en las distribuciones ordenadas por centralidad de intermediación es bastante menor (representando una mayor importancia) que su rango en las distribuciones ordenadas por centralidad de grado o por centralidad de cercanía.

Subred de enfermedad cardiovascular

Como podemos apreciar en la figura 3, con el ejemplo que alude al caso de las enfermedades cardiovasculares (nodos en color rojo), obtuvimos una subred con 205 nodos y 518 enlaces, que incluye información tanto para término femenino (female) (CG 17) como para el masculino (male) (CG 18) en los que no se aprecia una tendencia de conexión particular entre la enfermedad y el sexo.

Por otro lado, se aprecia un patrón de conexión indirecta y cada vez con una CG menor entre los nodos rojos y los nodos de color azul claro, azul oscuro, amarillo y verde, los cuales aluden a factores como la infancia, adolescencia, expectativa de vida, clase social, religión, derechos de las mujeres, derechos humanos, factores socioeconómicos, acceso a servicios de salud, acceso a servicios de salud de mujeres, estigma social, justicia social, características residenciales, políticas de salud, medicina de precisión, investigación biomédica y difusión de innovación.

Esta forma de relación indirecta sugiere que puede existir una brecha de información entre la información sobre la ECV y otros factores que pueden incidir o modificar su manejo médico, como la etiología y diagnóstico de dolor de pecho, la

relación médico paciente, la participación del paciente pertinente a su salud, cuidados postquirúrgicos, manejo de la enfermedad, mortalidad hospitalaria, emergencias, pronóstico, tasa de supervivencia, prevención secundaria, cuidados preoperatorios, complicaciones postoperatorias y efectos adversos.

Red emergente de la diferencia entre mujeres y hombres

Del análisis de las subredes semánticas se encontró que la red de mujeres estuvo constituida por 277 nodos y 3154 enlaces, y la de hombres por 237 nodos y 2834 enlaces. Como resultado de la diferencia de las subredes, se identificaron los términos que se encontraban en la red de mujeres y no en la de hombres. Se trató de una estructura con seis componentes, compuestos de 40 nodos y 135 enlaces (figura 4).

El componente que se aprecian en color azul y con CG 9, incluye algunos términos relacionados con las diferencias culturales, los migrantes y las disparidades en el cuidado de la salud. En rojo con CG 8, los derechos de las mujeres, la política pública y también términos sobre disparidades en salud. En amarillo con CG 7, los servicios de salud para mujeres y la circuncisión femenina. En morado con CG 5, aquellos términos relacionados con África, síndrome de inmunodeficiencia adquirida, y algunos componentes del sistema nacional de atención médica y la accesibilidad a los mismos.

En verde con CG 4, se aprecian términos relacionados con los centros de salud que brindan terapia y/o rehabilitación a personas dependientes de sustancias, México, y disparidades en el cuidado de la salud. Finalmente, en cian con CG 1 solo hay dos términos Nepal, y prolapso uterino. Por otro lado, el último grupo de

nodos de color cian, aparenta ser de poca importancia en la literatura por su grado de conexión tan pequeño, CG 1.

Modularidad

En la figura 5 se muestran los nodos intraconectados representativos de los MeSH, se aprecian las comunidades con aquellos términos que se relacionan principalmente con factores de riesgo. Los enlaces y su grosor simbolizan el flujo de información (de coocurrencia) intranodos. Las características topológicas de las subredes, correspondiente a la partición modular en comunidades se muestran en el cuadro 4.

El componente mayor conectado apreciado en figura 5a es *edad mediana*, tiene un IPR de 29% y un flujo de información (FI) de 12% (es de hacer notar que entre más grande el FI de un módulo, mayor es la coherencia interna de sus nodos entre sí y que la suma de la información total en las redes implica $FI=1.0$). Cuenta con 62 nodos, éstos representan los términos más relacionados con la edad y los factores socioeconómicos, accesibilidad a los servicios de salud, patrones de práctica médica, y otros conceptos relacionados con la planificación para la asignación equitativa de los recursos de salud disponibles. Es interesante notar que los patrones de conectividad de los nodos que forman parte de una comunidad son característicos. Con el fin de hacerlos explícitos hemos determinado un conjunto de medidas de centralidad y otros atributos topológicos para los nodos en el contexto de sus respectivas comunidades (La tabla correspondiente a la lista de atributos topológicos de la comunidad 5a se ha incluido en el cuadro suplementario 2).

En la figura 5b, se muestra la comunidad asociada con la condición de ser mujer (IPR 12% y FI 10%), algunos de los 19 términos más relacionados con éste son: disparidad en salud, esperanza de vida, atención médica de parto, calidad de la asistencia sanitaria, VIH, entre otros. (De manera similar al caso anterior, la tabla correspondiente a la lista de atributos topológicos de la comunidad 5b se ha incluido en el cuadro suplementario 3).

En la figura 5c, se muestra el módulo asociado con los factores de riesgo (IPR 7.9% y FI 3.3%) y 18 conceptos relacionados con la justicia social, la identidad de género, los derechos de las mujeres, el paternalismo, la cultura, entre otros.

Resulta interesante distinguir el relativamente bajo papel que estos temas han jugado en la literatura biomédica analizada. (La tabla correspondiente a la lista de atributos topológicos de la comunidad 5c se ha incluido en el cuadro suplementario 4).

Discusión

Minería computarizada de términos MeSH

Consistente con lo que se ha reportado en la literatura, la minería computarizada como la que se realizó en este trabajo, ofrece métodos poderosos, rápidos y de bajo costo para la exploración masiva de textos científicos, Utilizar la ontología MeSH en lugar del texto completo no solo reduce el tiempo de cálculo matemático en algoritmos de coocurrencia, sino que también permite un mayor rendimiento en el análisis del conjunto de datos. Por lo tanto, se pueden extraer características representativas de grandes cantidades de literatura con precisión utilizando estos términos MeSH de alta calidad si se exploran adecuadamente (36).

El procesamiento de un corpus de texto simple a un lenguaje natural apoya el diseño de mecanismos eficaces de comunicación computacional para ser interpretados por programas que simulen el contenido semántico de los conceptos y favorezca la construcción de mapas de conectividad ontológica (37). Para el caso que nos ocupa sobre disparidades en salud, utilizamos la construcción de redes semánticas para inferir cómo se reporta la información en algunos documentos científicos.

Construcción de redes y subredes semánticas

Como se mencionó en la introducción, las posibles causas de las disparidades son multifactoriales (38). Por ejemplo, en la enfermedad cardiovascular, si bien estas disparidades se han asociado con la atención médica y el manejo clínico existen otros elementos que interactúan y pueden modificar la expresión de la enfermedad (39-42). Así mismo, se ha reportado que el papel de las estructuras socioeconómicas y los sistemas sanitarios han sido responsables de la presencia de algunas disparidades por sexo en varios sentidos: al mostrar apatía para incidir y controlar en factores de riesgo modificables, indiferencia en el reconocimiento de signos y síntomas, insensibilidad en el acceso y uso de tratamientos químicos o quirúrgicos de última generación, o el desinterés por ayudar a la recuperación, pero sobre todo el rezago en generación de evidencia basada en el sexo para impulsar las decisiones clínicas y desarrollar políticas necesarias para mejorar la calidad y eficiencia de la atención cardiovascular de las mujeres (43,44).

Lo que observamos en la red principal de disparidades en salud y sexismo muestra que los MeSH que aparecen con mayor frecuencia, entre los que hacen referencia a condiciones médicas, están relacionados con la ECV. Son 17 nodos,

entre ellos se encontraron los siguientes MeSH: *Cardiac Surgical Procedures, Myocardial Infarction, Acute Coronary Syndrome, Coronary Angiography, Coronary Artery Disease, Fibrinolytic Agents, Myocardial Revascularization, Percutaneous Coronary Intervention, Thrombolytic Therapy.*

Estos hallazgos eran esperados dado el incremento de la publicación científica relacionada con la ECV y el sexo femenino, sobre todo porque la enfermedad coronaria, tradicionalmente considerada una enfermedad masculina, se ha convertido en una gran amenaza biológica, hormonal y genética para las mujeres (45,46).

En las últimas dos décadas, la investigación centrada en mujeres con riesgo de enfermedad isquémica del corazón ha ayudado a aclarar nuestra comprensión de algunos de los factores específicos del sexo que son importantes en la prevención y detección temprana de disparidades en salud, particularmente en aquellas mujeres que están en desventaja social debido a su nivel de ingresos, nivel educativo, factores ambientales y residenciales (43).

Por otro lado, la disparidad en la mortalidad por ECV entre mujeres y hombres también impulsó la investigación específica por sexo, así como, el desarrollo de directrices específicas y el inicio de campañas de sensibilización en Estados Unidos, de esta manera se ha reportado una sensible disminución del 39% en las tasas de mortalidad por ECV en las mujeres de 1999 a 2011. Sin embargo, a pesar de esta disminución, la carga de la enfermedad, la coexistencia con otras enfermedades, así como la prevalencia de los factores de riesgo siguen siendo altas (43).

Con el ejemplo de la sub red de MeSH de ECV podemos percibir una aparente disonancia entre los MeSH relacionados con la enfermedad y los términos para describir los determinantes sociales, factores de riesgo, edad, actividad física y algunas comorbilidades. Este hecho nos hace pensar que la información sobre disparidades en salud, entre los factores biomédicos y el contexto social, continúa sin resolverse al menos en materia de evidencia científica (43,47).

Como se puede apreciar tanto en la red principal como en la subred de ECV, coexisten términos asociados con otras enfermedades crónicas que también parecen relacionados con las disparidades de sexo. Esto también es consistente con una creciente evidencia científica que sugiere que el sexo es un factor independiente que influye en la prevalencia de enfermedades como cáncer, enfermedades mentales, VIH y SIDA, así como en los efectos de los medicamentos, los eventos adversos y el manejo del dolor (48,49).

En cuanto a los factores de riesgo, la investigación ha demostrado que aunque los hombres y las mujeres comparten algunas características que aumentan su probabilidad de sufrir una enfermedad coronaria (tabaquismo, colesterol elevado, diabetes, estilo de vida sedentario, hipertensión, obesidad), varias de estas causas son más agudas en las mujeres (50). Además, ciertos factores de riesgo específicos del sexo, como el inicio temprano de la menopausia, enfermedades inflamatorias como el lupus y la artritis reumatoide, y las complicaciones del embarazo como la preeclampsia y la diabetes gestacional se asocian con una mayor incidencia de enfermedades del corazón (51). Hasta un cincuenta por ciento de la reducción de la mortalidad por enfermedades cardiovasculares se puede atribuir a la reducción de estos factores de riesgo principales, y se necesita

un enfoque continuo en la sensibilización y la detección temprana para seguir reduciendo las tasas de mortalidad (43).

Los hallazgos respecto a la red emergente, los términos que solo se encontraron en la subred de mujeres y no en la de hombres confirma lo que se ha reportado en la literatura, que la salud y el bienestar de las mujeres están conformados por una combinación de políticas de atención médica que afectan el tipo de cobertura de seguro de salud a la que pueden tener acceso a servicios preventivos, de detección y tratamiento (52-55). Estos hallazgos armonizan hasta cierto punto, con la teoría del modelo biopsicosocial sobre las enfermedades crónicas, causadas y mantenidas no solo por factores biológicos, sino también por condiciones psicológicas y sociales (56).

Por otro lado, aunque los nodos de color cian de la red emergente aparentan ser de poca importancia, el problema no parece ser menor; la organización de derechos humanos Amnistía Internacional ha señalado que alrededor del 10% de las mujeres en Nepal ha experimentado un prolapso uterino debido a una discriminación de sexo. Además, se dice que estas mujeres enfrentan un fuerte estigma social y son señaladas como perezosas por estar demasiado enfermas para trabajar, en pocas palabras, otra forma de disparidad en el cuidado de la salud (57).

Modularidad

Respecto al análisis topológico de la modularidad utilizando el algoritmo Infomap, éste nos permitió organizar muchos nodos de acuerdo con su distribución estructural en cada comunidad. Si bien la comunidad mayormente conectada, de *edad mediana* (IPR=29%, FI=12%), es la que contiene dos de los principales

términos relacionados con los factores de riesgo (edad y socioeconómicos), los términos relacionados con la calidad de la atención se hacen más contundentes con este análisis en el resto de las comunidades, por ejemplo los relacionados con la asignación equitativa de los recursos de salud disponibles, la justicia social y los derechos de las mujeres. Es de hacer notar que este fenómeno no solo se presenta asociado a los padecimientos cardiacos, sino también a otras comorbilidades como lo notamos en la red principal.

Estos resultados son consistentes con lo que se ha informado en la literatura sobre los factores que influyen o modifican la calidad de la atención y las disparidades por sexo, entre los que se incluyen la mala comunicación paciente-médico, la cultura, las opiniones holísticas de la mente y el cuerpo, la priorización del manejo de los síntomas versus la cura de la enfermedad y la participación de la familia en la toma de decisiones médicas (52). Así mismo, los factores tanto internos como externos al sistema de prestación de servicios de salud como, la falta de acceso a la atención médica, la falta de atención oportuna y la escasa protección financiera, también pueden incidir en la calidad de la atención (53,54). En cuanto a la comunidad sobre la condición de ser mujer, también se nota una relación con términos sobre la calidad asistencial y el VIH, por ejemplo. Desde la aparición del VIH/SIDA, la discriminación desde los sistemas de salud ha permanecido en el centro de las experiencias negativas de las personas que viven con el virus VIH (58). Por otro lado, sabemos que las mujeres experimentan diferentes síntomas clínicos y complicaciones debido al VIH. Sin embargo, una revisión de 2016 sobre la inclusión de mujeres en las investigaciones sobre el VIH en los EE. UU. encontró que las mujeres representaban solo el 19.2% de los

participantes en los estudios antirretrovirales, el 38.1% en los estudios de vacunación y el 11.1% en estudios para encontrar cura (59,60). Nuevamente otro ejemplo de disparidad en el cuidado de la salud asociado al sexismo, y que, para fines de este artículo, se vuelve más contundente gracias al análisis de comunidades de términos MeSH.

Por lo tanto, si se quiere contribuir a la corrección de las disparidades en salud se deben explorar integralmente los mecanismos sociales, biológicos y clínicos de los hombres y de las mujeres. Aunque las diferencias sexuales están cada vez mejor documentadas en la literatura biomédica, como evidencian casos como la prevalencia de los factores de riesgo cardiovascular, las condiciones fisiológicas y psicosociales, así como las manifestaciones clínicas y la incidencia de la enfermedad, los médicos se han referido a la atención médica de la mujer como *medicina de bikini* y han asumido que pueden diagnosticar y tratar a ambos sexos de la misma manera (61). Esta tendencia también se percibe como una manera de perpetuar la disparidad en la calidad de la atención médica.

Acerca del análisis de la literatura asistido por computadora

Dada la abrumadora cantidad de información disponible en las bases de datos biomédicas, contar con un recurso que pueda leer automáticamente la información y generar patrones de visualización de conceptos y asociaciones es invaluable, especialmente para los tomadores de decisiones. Por lo tanto, en términos de metodología de investigación, este artículo demuestra la utilidad de un novedoso método basado en una búsqueda rápida y masiva de evidencia científica mediante redes semánticas de términos MeSH.

En la práctica, la identificación de módulos o comunidades es fundamental para inferir relaciones entre los MeSH a partir de la topología de redes semánticas y en este caso, puede ser una guía confiable para futuros estudios sobre disparidades en salud y sexismo, así como para profundizar en sus implicaciones teóricas.

Como hemos señalado, nuestros análisis muestran algunos aspectos esenciales de el fenómeno en cuestión que aparentemente han sido pasados por alto de manera consistente, lo que ha hecho aún más profunda la brecha que hay que cerrar y pasar de la discusión de estas desigualdades a la formulación de políticas de salud pública para atenderlas desde un enfoque social centrado en el individuo. Por tanto podemos señalar que la investigación que examina, tanto la información para ambos sexos, como los contextos sociales amplía la generalización de los resultados y la aplicabilidad al escenario médico-clínico (11).

Hacer lo contrario puede influir en interpretaciones erróneas en la toma de decisiones y comportamientos sociales ante el proceso salud enfermedad, o tener consecuencias adversas en la provisión de servicios sanitarios, así como, un posible perjuicio de la formulación de políticas de salud, y en algunos casos, hasta dificultar la lucha contra posibles estructuras de poder, incluidas las jerarquías de conocimiento, que perpetúan la inequidad y la mala salud (6,9,12).

Conflicto de intereses

Los autores declaran no tener ningún conflicto de interés.

Financiación

La presente investigación no ha recibido ayudas específicas provenientes de agencias del sector público, sector comercial o entidades con ánimo de lucro.

Referencias

1. **Payne S.** The health of women and girls: how can we address gender equality and gender equity?. *Semin Reprod Med.* 2015;33:53-60.
<https://doi.org/10.1055/s-0034-1395280>
2. **Davies SE, Harman S, Manjoo R, Tanyag M, Wenham C.** Why it must be a feminist global health agenda. *Lancet.* 2019;393:601-3.
[https://doi.org/10.1016/S0140-6736\(18\)32472-3](https://doi.org/10.1016/S0140-6736(18)32472-3)
3. **Schopper M, Irnich D.** Gender and its implications for cardiothoracic perioperative care and anesthesia. *Thorac Cardiovasc Surg.* 2013;61:7-14.
<https://doi.org/10.1055/s-0032-1331038>
4. **Glymour MM, Spiegelman D.** Evaluating public health interventions: 5. causal inference in public health research-do sex, race, and biological factors cause health outcomes? *Am J Public Health.* 2017;107:81-5.
<https://doi.org/10.2105/AJPH.2016.303539>
5. **Stewart M, Kushner KE, Gray J, Hart DA.** Promoting gender equity through health research: impacts and insights from a Canadian initiative. *Glob Health Promot.* 2013;20:25-38.
<https://doi.org/10.1177/1757975913476903>
6. **Heidari S, Babor TF, De Castro P, Tort S, Curno M.** Sex and Gender Equity in Research: rationale for the SAGER guidelines and recommended use. *Res Integr Peer Rev.* 2016;1:2. <https://doi.org/10.1186/s41073-016-0007-6>

7. **Potterat MM, Monnin Y, Pechère A, Guessous I.** Women, forgotten by clinical research. *Rev Med Suisse.* 2015;11:1733-6.
8. **Hamberg K.** Gender bias in medicine. *Womens Health.* 2008;4:237-43.
<https://doi.org/10.2217/17455057.4.3.237>
9. **Sugimoto CR, Ahn YY, Smith E, Macaluso B, Larivière V.** Factors affecting sex-related reporting in medical research: a cross-disciplinary bibliometric analysis. *Lancet.* 2019;393:550-9.
[https://doi.org/10.1016/S0140-6736\(18\)32995-7](https://doi.org/10.1016/S0140-6736(18)32995-7)
10. **Ward PR, Meyer SB, Verity F, Gill TK, Luong TC.** Complex problems require complex solutions: the utility of social quality theory for addressing the Social Determinants of Health. *BMC Public Health.* 2011;11:630.
<https://doi.org/10.1186/1471-2458-11-630>
11. **Diderichsen F, Evans T, Whitehead M.** Bases sociales de las disparidades en salud. En: Evans T, editor. *Desafío a la falta de equidad en la salud.* Washington: OPS/Fundación Rockefeller; 2002.p.13-26,185-205.
12. **Sen G, George A, Östlin P.** The case for gender equity in health research. *J Health Manag.* 2002;4:99-117.
13. **Sakaguchi AM.** Health equities worldwide. *J Health Polit Policy Law.* 2003;28:552-6.
14. **Graham H.** *Unequal lives: health and socioeconomic inequalities.* Maidenhead: McGraw-Hill Education; 2007. p.82-113.
15. **Golden TL, Wendel ML.** Public health's next step in advancing equity: re-evaluating epistemological assumptions to move social determinants from

theory to practice. *Front Public Health*. 2020;8:131.

<https://doi.org/10.3389/fpubh.2020.00131>

16. **Piccini P, Montagnani C, de Martino M.** Gender disparity in pediatrics: a review of the current literature. *Ital J Pediatr*. 2018;44:1-6
<https://doi.org/10.1186/s13052-017-0437-x>
17. **Martínez-García M, Salinas-Ortega M, Estrada-Arriaga I, Hernández-Lemus E, García-Herrera R, Vallejo M.** A systematic approach to analyze the social determinants of cardiovascular disease. *PLoS One*. 2018;13:e0190960. <https://doi.org/10.1371/journal.pone.0190960>
18. **Martínez-García M, Vallejo M, Hernández-Lemus E, Álvarez-Díaz JA.** Novel methods of qualitative analysis for health policy research. *Health Res Policy Syst*. 2019;17:6. <https://doi.org/10.1186/s12961-018-0404-z>
19. **Perl Y, Chen Z, Halper M, Geller J, Zhang L, Peng Y.** The cohesive metaschema: a higher-level abstraction of the UMLS Semantic Network. *J Biomed Inform*. 2002;35:194-212. [https://doi.org/10.1016/s1532-0464\(02\)00528-2](https://doi.org/10.1016/s1532-0464(02)00528-2)
20. **Siqueiros-García JM, Hernández-Lemus E, García-Herrera R, Robina-Galatas A.** Mapping the structure and dynamics of genomics-related MeSH terms complex networks. *PLoS One*. 2014;9:e92639.
<https://doi.org/10.1371/journal.pone.0092639>
21. **Barabási AL.** *Network science*. Cambridge, UK: Cambridge University Press; 2016. p. 45-50.
22. **Tang L, Bie B, Zhi D.** Tweeting about measles during stages of an outbreak: A semantic network approach to the framing of an emerging

infectious disease. *Am J Infect Control*. 2018;46:1375-80.

<https://doi.org/10.1016/j.ajic.2018.05.019>

23. **Utsumi A.** A complex network approach to distributional semantic models. *PLoS One*. 2015;10:e0136277.

<https://doi.org/10.1371/journal.pone.0136277>

24. **Theodosiou T, Vizirianakis IS, Angelis L, Tsiftaris A, Darzentas N.**

MeSHy: Mining unanticipated PubMed information using frequencies of occurrences and concurrences of MeSH terms. *J Biomed Inform*.

2011;44:919-926. <https://doi.org/10.1016/j.jbi.2011.05.009>

25. **Yang H, Lee HJ.** Research Trend Visualization by MeSH Terms from PubMed. *Int J Environ Res Public Health*. 2018;15:1113.

<https://doi.org/10.3390/ijerph15061113>

26. **Zhang H, Fiszman M, Shin D, Miller CM, Roseblat G, Rindfleisch TC.**

Degree centrality for semantic abstraction summarization of therapeutic studies. *J Biomed Inform*. 2011;44:830-8.

<https://doi.org/10.1016/j.jbi.2011.05.001>

27. **Ruiz JB, Barnett GA.** Exploring the presentation of HPV information online:

A semantic network analysis of websites. *Vaccine*. 2015;33:3354-9.

<https://doi.org/10.1016/j.vaccine.2015.05.017>

28. **Zemla JC, Austerweil JL.** Modeling semantic fluency data as search on a semantic network. *Cogsci*. 2017;2017:3646-51.

29. **Kohl M, Wiese S, Warscheid B.** Cytoscape: software for visualization and analysis of biological networks. *Methods Mol Biol.* 2011;696:291-303.
https://doi.org/10.1007/978-1-60761-987-1_18
30. **Alcalá-Corona SA, Espinal-Enríquez J, de Anda-Jáuregui G, Hernández-Lemus E.** The Hierarchical Modular Structure of HER2+ Breast Cancer Network. *Front Physiol.* 2018;9:1423.
<https://doi.org/10.3389/fphys.2018.01423>
31. **Alcalá-Corona SA, de Anda-Jáuregui G, Espinal-Enríquez J, Tovar H, Hernández-Lemus E.** Network Modularity and Hierarchical Structure in Breast Cancer Molecular Subtypes. En: Morales A, Gershenson C, Braha D, Minai A, Bar-Yam Y, editores. *Unifying Themes in Complex Systems IX.* Springer Proceedings in Complexity. Cambridge: Springer International Publishing; 2018.
32. **Riolo MA, Newman MEJ.** Consistency of community structure in complex networks. *Physical Review E.* 2020;101.5:052306.
33. **Rosvall M, Bergstrom CT.** Maps of random walks on complex networks reveal community structure. *Proc Natl Acad Sci USA.* 2008;105:1118-23.
<https://doi.org/10.1073/pnas.0706851105>
34. **Bohlin L, Edler D, Lancichinetti A, Rosvall M.** Community Detection and Visualization of Networks with the Map Equation Framework. En: Ding Y, Rousseau R, Wolfram D, editores. *Measuring Scholarly Impact.* Cambridge: Springer International Publishing; 2014.
35. **Rosvall M, Axelsson D, Bergstrom CT.** The map equation. *Eur Phys J Spec Top.* 2009;178:13-23. <https://doi.org/10.1140/epjst/e2010-01179-1>

36. **Renganathan V.** Text mining in biomedical domain with emphasis on document clustering. *Healthc Inform Res.* 2017;23:141-6.
<https://doi.org/10.4258/hir.2017.23.3.141>
37. **Simon C, Davidsen K, Hansen C, Seymour E, Barnkob MB, Olsen LR.** BioReader: a text mining tool for performing classification of biomedical literature. *BMC Bioinformatics.* 2019;19(Suppl 13):57.
<https://doi.org/10.1186/s12859-019-2607-x>
38. **Araújo C, Pereira M, Laszczyńska O, Dias P, Azevedo A.** Sex-related inequalities in management of patients with acute coronary syndrome- results from the EURHOBOP study. *Int J Clin Pract.* 2018;72:
<https://doi.org/10.1111/ijcp.13049>
39. **Nestler DM, Gilani WI, Anderson RT, Bellolio MF, Branda ME, LeBlanc A, et al.** Does gender bias in cardiac stress testing still exist? A videographic analysis nested in a randomized controlled trial. *Am J Emerg Med.* 2017;35:29-35. <https://doi.org/10.1016/j.ajem.2016.09.054>
40. **Jortveit J, Govatsmark RE, Langørgen J, Hole T, Mannsverk J, Olsen S, et al.** Gender differences in the assessment and treatment of myocardial infarction. *Tidsskr Nor Laegeforen.* 2016;136:1215-22.
<https://doi.org/10.4045/tidsskr.16.0224>
41. **Aaberge L.** Equal treatment for myocardial infarction patients?. Likeverdig behandling ved hjerteinfarkt?. *Tidsskr Nor Laegeforen.* 2016;136:1181.
<https://doi.org/10.4045/tidsskr.16.0532>
42. **Leurent G, Garlantézec R, Auffret V, Hacot JP, Coudert I, Filippi E, et al.** Gender differences in presentation, management and inhospital outcome in

- patients with ST-segment elevation myocardial infarction: data from 5000 patients included in the ORBI prospective French regional registry. Arch Cardiovasc Dis. 2014;107:291-8. <https://doi.org/10.1016/j.acvd.2014.04.005>
43. **Rosen SE, Henry S, Bond R, Pearte C, Mieres JH.** Sex-specific disparities in risk factors for coronary heart disease. Curr Atheroscler Rep. 2015;17:49. <https://doi.org/10.1007/s11883-015-0523-8>
44. **Woodward M.** Cardiovascular disease and the female disadvantage. Int J Environ Res Public Health. 2019;16:1165. <https://doi.org/10.3390/ijerph16071165>
45. **Humphries K, Izadnegahdar M, Sedlak T, Saw J, Johnston N, Schenck-Gustafsson K, et al.** Sex differences in cardiovascular disease-Impact on care and outcomes. Front Neuroendocrinol. 2017;46:46-70. <https://doi.org/10.1016/j.yfrne.2017.04.001>
46. **Kane AE, Howlett SE.** Differences in cardiovascular aging in men and women. Adv Exp Med Biol. 2018;1065:389-411. https://doi.org/10.1007/978-3-319-77932-4_25.
47. **Lundberg GP, Mehta LS, Sanghani RM, Patel HN, Aggarwal NR, Aggarwalet NT, et al.** Heart centers for women: historical perspective on formation and future strategies to reduce cardiovascular disease. Circulation. 2018;138:1155-65. <https://doi.org/10.1161/CIRCULATIONAHA.118.035351>
48. **Goss PE, Strasser-Weippl K, Lee-Bychkovsky BL, Fan L, Li J, Chavarri-Guerra Y, et al.** Challenges to effective cancer control in China, India, and

- Russia. *Lancet Oncol.* 2014;15:489-538. [https://doi.org/10.1016/S1470-2045\(14\)70029-4](https://doi.org/10.1016/S1470-2045(14)70029-4)
49. **Thiels CA, Holst KA, Ubl DS, McKenzie TJ, Zielinski MD, Farley DR, et al.** Gender disparities in the utilization of laparoscopic groin hernia repair. *J Surg Res.* 2017;210:59-68. <https://doi.org/10.1016/j.jss.2016.10.028>
50. **Appelman Y, Van Rijn BB, Monique E, Boersma E, Peters SA.** Sex differences in cardiovascular risk factors and disease prevention. *Atherosclerosis.* 2015;241:211-8. <https://doi.org/10.1016/j.atherosclerosis.2015.01.027>
51. **Young L, Cho L.** Unique cardiovascular risk factors in women. *Heart.* 2019;105:1656-60. <https://doi.org/10.1136/heartjnl-2018-314268>
52. **Mensah GA.** Health and health care disparities in cancer and cardiovascular diseases - learning from one to improve the other. *Ethn Dis.* 2018;28:133-4. <https://doi.org/10.18865/ed.28.3.133>
53. **Wasserman J, Palmer RC, Gomez MM, Berzon R, Ibrahim SA, Ayanian JZ.** Advancing health services research to eliminate health care disparities. *Am J Public Health.* 2019;109(Supl.1):S64-9. <https://doi.org/10.2105/AJPH.2018.304922>
54. **Martínez-García M, Vargas-Barrón J, Bañuelos-Téllez F, González-Pacheco H, Fresno C, Hernández-Lemus E, et al.** Public insurance program impact on catastrophic health expenditure on acute myocardial infarction. *Public Health.* 2018;158:47-54. <https://doi.org/10.1016/j.puhe.2018.01.025>

55. **Brindis CD, Freund KM.** The ramifications of recent health policy actions for cardiovascular care of women: Progress, threats, and opportunities. *Clin Cardiol.* 2018;41:173-8. <https://doi.org/10.1002/clc.22896>
56. **Fernández-López JA.** Towards a better understanding of patients. *Semergen.* 2013;39,8:403-5.
<https://doi.org/10.1016/j.semerg.2013.06.007>
57. **Gulland A.** Gender discrimination is blamed for high rates of uterine prolapse in Nepal. *BMJ.* 2014;348. <https://doi.org/10.1136/bmj.g1756>
58. **Castilho JL, Escuder MM, Veloso V, Gomes JO, Jayathilake K, Ribeiro S, et al.** Trends and predictors of non-communicable disease multimorbidity among adults living with HIV and receiving antiretroviral therapy in Brazil. *J Int AIDS Soc.* 2019;22:e25233. <https://doi.org/10.1002/jia2.25233>
59. **Curno MJ, Rossi S, Hodges-Mameletzis I, Johnston R, Price MA, Heidari S.** A systematic review of the inclusion (or exclusion) of women in HIV research: from clinical studies of antiretrovirals and vaccines to cure strategies. *J Acquir Immune Defic Syndr.* 2016;71:181-8.
<https://doi.org/10.1097/QAI.0000000000000842>
60. **Perez CC.** *Invisible Women: Exposing Data Bias in a World Designed for Men.* New York: Random House; 2019.
61. **Ouyang P, Wenger NK, Taylor D, Rich-Edwards JW, Steiner M, Shaw LJ, et al.** Strategies and methods to study female-specific cardiovascular health and disease: a guide for clinical scientists. *Biol Sex Differ.* 2016;7:19.
<https://doi.org/10.1186/s13293-016-0073-y>

Figuras

Figura 1.

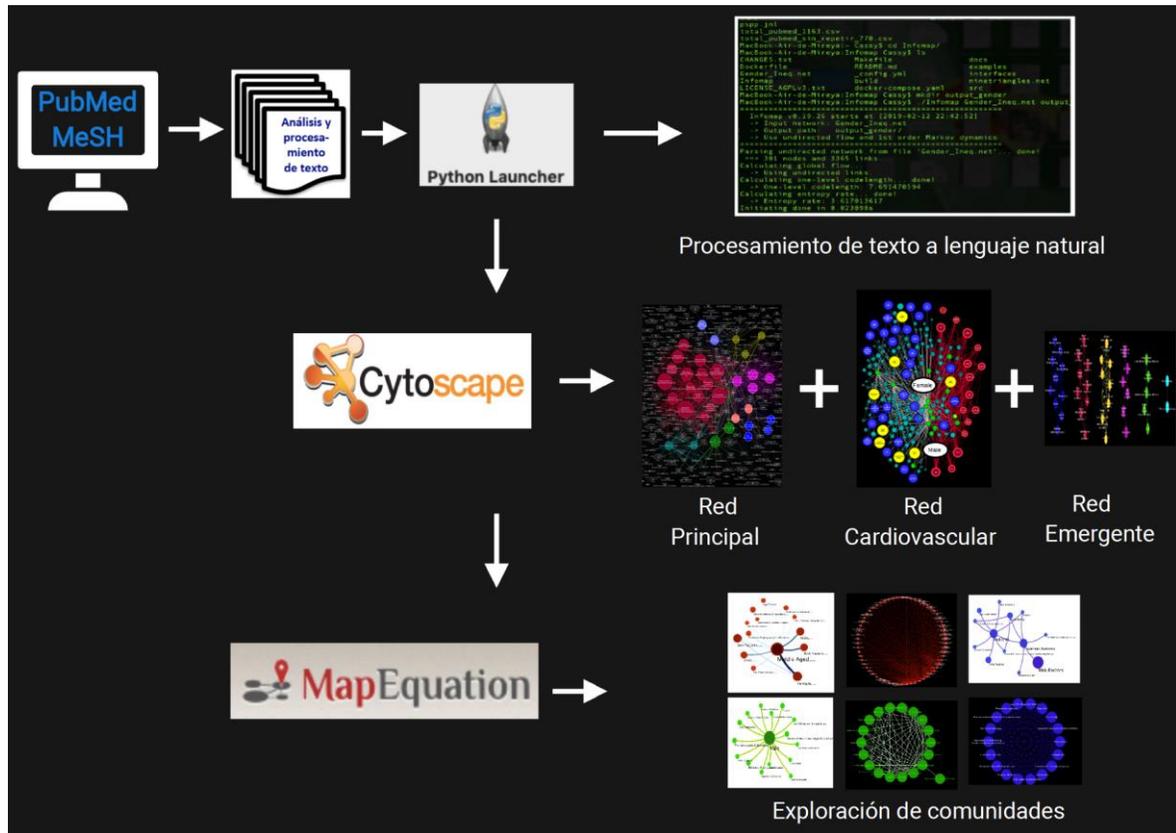


Figura 1. Flujo de trabajo. Una vez que se realizó la minería computarizada de la literatura, se procesó el corpus de información, de texto a lenguaje natural. Posteriormente, se construyeron los mapas de conectividad y las redes semánticas. Finalmente, se aplicó un algoritmo de agrupación para detectar módulos o comunidades y completar el análisis.

Figura 2

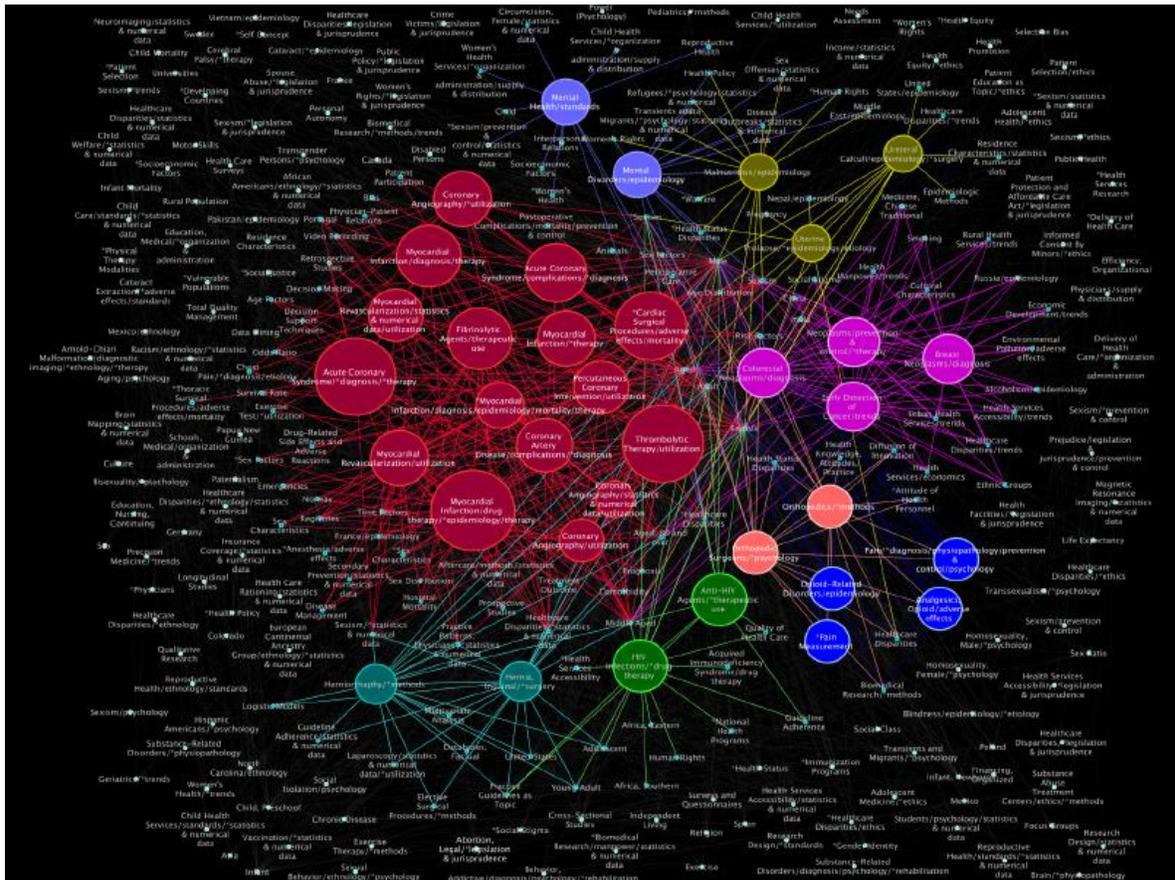


Figura 2. Red principal. Esta figura resalta los nodos relacionados con algunas enfermedades crónicas, éstos se aumentaron de tamaño y se pintaron de colores brillantes para su mejor distinción.

Figura 3

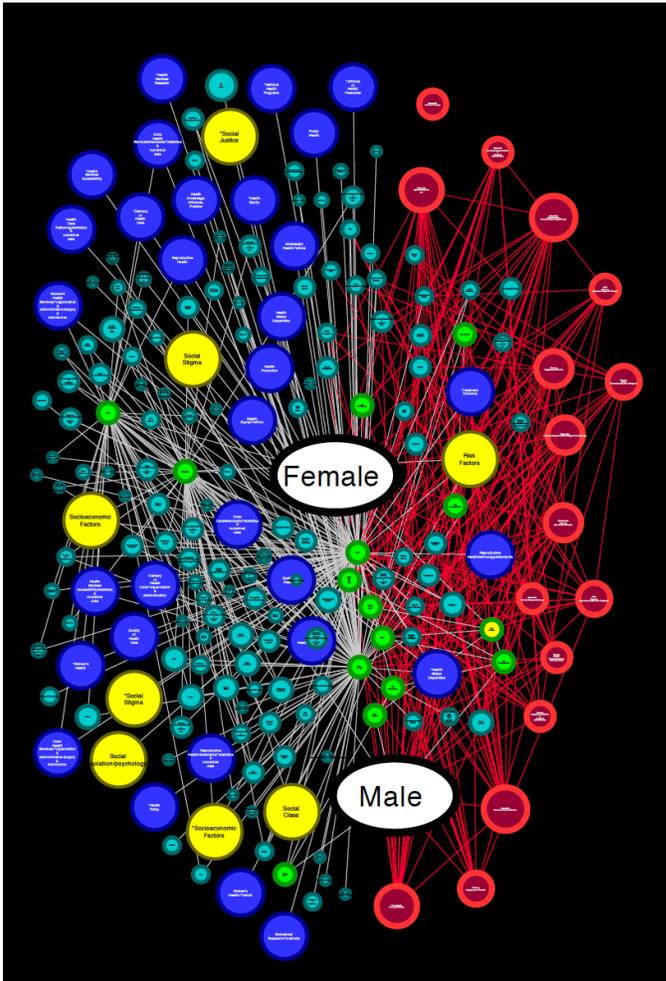


Figura 3. Red de enfermedades cardiacas. En esta red se aprecia la agrupación por nodos de colores ordenados para mostrar la conexión indirecta de los términos MeSH relacionados con los padecimientos cardiacos (en rojo) y aquellos otros que hacen referencia a términos de factores físicos, del sistema de salud, determinantes sociales y otros (en colores azul claro, azul oscuro, amarillo y verde).

Figura 4

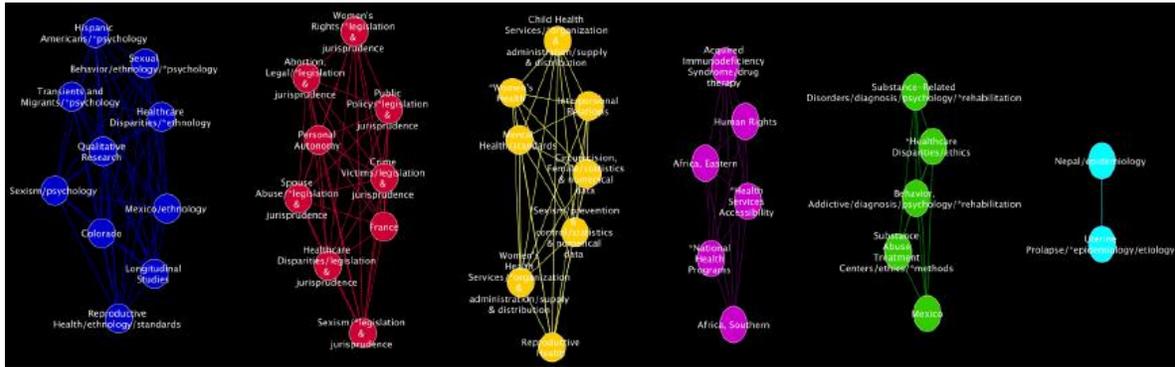


Figura 4. Red emergente. Esta red agrupa los nodos por colores y CG: centralidad de grado. Fue reconstruida a partir de la diferencia de nodos y vecinos de las subredes de hombres y mujeres, para conocer los términos incluidos en una y no en la otra, los cuales se aprecian en esta figura.

Figura 5

Visualización de las principales comunidades y sus respectivas redes de
Disparidades sanitarias y sexismo

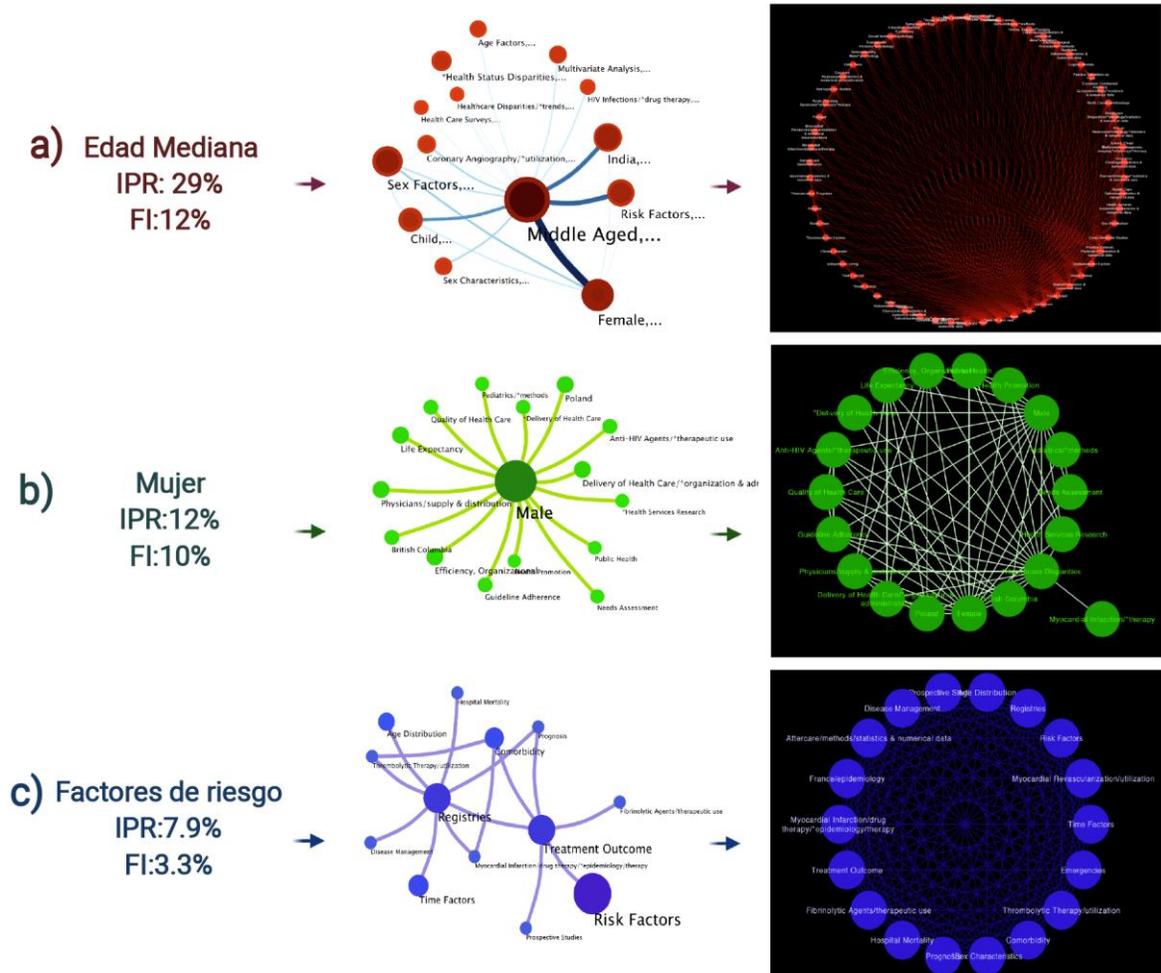


Figura 5. Visualización modular de las principales comunidades y sus respectivas redes. Los círculos más grandes representa las comunidades que comparten más enlaces de información entre nodos. El nombre de las etiquetas de las comunidades está dado por el nodo con el mayor PageRank.

Cuadros

Cuadro 1. Características topológicas de las redes analizadas.

Red	Nodos	Aristas	Coefficiente de agrupamiento	Longitud de camino	Centralidad	Número promedio de vecinos	Densidad de red
Principal	301	3365	0.910	1.962	0.851	22.359	0.075
EVC	205	518	0.194	2.766	0.480	5.054	0.025
Mujeres	277	3154	0.907	1.917	0.924	22.773	0.083
Hombres	237	2834	0.902	1.899	0.906	23.916	0.101
Emergente	40	135	0.95	1.0	0.061	6.75	0.173
EVC: Enfermedad cardiovascular							

Cuadro 2. Comparación entre diferentes medidas de centralidad para los nodos asociados a condiciones médicas relacionados con la enfermedad cardiovascular.

Términos MeSH	CVP	CG	PCMC	CI	CC	CA
Fibrinolytic Agents/therapeutic use	0.056516256	23	1.92857143	0	0.51851852	1
Myocardial Infarction/drug therapy/*epidemiology/therapy	0.056516256	23	1.92857143	0	0.51851852	1
Myocardial Revascularization/utilization	0.056516256	23	1.92857143	0	0.51851852	1
Thrombolytic Therapy/utilization	0.056516256	23	1.92857143	0	0.51851852	1
*Myocardial Infarction/diagnosis/epidemiology/mortality/therapy	0.04928961	18	1.93877551	0	0.51578947	1
Coronary Angiography/utilization	0.049289722	18	1.93877551	0	0.51578947	1
Percutaneous Coronary Intervention/utilization	0.04928957	18	1.93877551	0	0.51578947	1
Acute Coronary Syndrome/complications/*diagnosis	0.04531952	17	1.97619048	0	0.5060241	1
Chest Pain/*diagnosis/etiology	0.04531952	17	1.97619048	0	0.5060241	1
Coronary Angiography/*utilization	0.04531952	17	1.97619048	0	0.5060241	1
Coronary Artery Disease/complications/*diagnosis	0.04531952	17	1.97619048	0	0.5060241	1
*Cardiac Surgical Procedures/adverse effects/mortality	0.036228206	15	1.95578231	0	0.51130435	1
Acute Coronary Syndrome/*diagnosis/*therapy	0.04339972	15	1.9829932	0	0.50428816	1
Coronary Angiography/statistics & numerical data/utilization	0.043399736	15	1.9829932	0	0.50428816	1
Myocardial Infarction/diagnosis/therapy	0.043399762	15	1.9829932	0	0.50428816	1
Myocardial Revascularization/statistics & numerical data/utilization	0.04339972	15	1.9829932	0	0.50428816	1
Myocardial Infarction/*therapy	0.007807993	2	2.38435374	0	0.41940086	1
CVP: Centralidad de valor propio. CG: Centralidad de Grado. PCMC: Promedio de caminos más cortos. CI: Centralidad de intermediación. CC: Centralidad de cercanía. CA: Coeficiente de agrupamiento.						

Cuadro 3. Comparación entre diferentes medidas de centralidad para los nodos asociados a condiciones médicas relacionados con otras enfermedades.

Términos MeSH	CVP	CG	PCMC	CI	CC	CA
Breast Neoplasms/diagnosis	0.06038379	26	1.91836735	0	0.5212766	1
Colorectal Neoplasms/diagnosis	0.060383793	26	1.91836735	0	0.5212766	1
Neoplasms/prevention & control/*therapy	0.06038379	26	1.91836735	0	0.5212766	1
Hernia, Inguinal/*surgery	0.05517667	20	1.96598639	0	0.50865052	1
Arnold-Chiari Malformation/diagnostic imaging/*ethnology/*therapy	0.044969346	19	2.10544218	0	0.47495961	1
HIV Infections/*drug therapy	0.042333294	19	1.94217687	1.51 E-04	0.51488616	0.74 269 006
*Pain Measurement	0.030480828	16	1.97278912	0	0.50689655	1
Opioid-Related Disorders/epidemiology	0.030480828	16	1.97278912	0	0.50689655	1
Orthopedics/*methods	0.030480828	16	1.97278912	0	0.50689655	1
Pain/*diagnosis/physiopathology/prevention & control/psychology	0.030480828	16	1.97278912	0	0.50689655	1
Malnutrition/epidemiology	0.029936733	15	1.95578231	0	0.51130435	1
Mental Disorders/epidemiology	0.029936733	15	1.95578231	0	0.51130435	1
Cerebral Palsy/*therapy	0.034478564	14	1.95238095	0	0.51219512	1
Acquired Immunodeficiency Syndrome/drug therapy	0.03175653	13	1.96258503	0	0.50953206	1
Blindness/epidemiology/*etiology	0.02833061	12	1.95918367	0	0.51041667	1
Cataract Extraction/*adverse effects/standards	0.02833061	12	1.95918367	0	0.51041667	1
Cataract/*epidemiology	0.02833061	12	1.95918367	0	0.51041667	1
Disabled Persons	0.015048198	12	2.33333333	0	0.42857143	1
Ureteral Calculi/epidemiology/*surgery	0.04004379	12	1.96598639	0	0.50865052	1
Substance-Related Disorders/*physiopathology	0.022431524	11	2.02380952	0	0.49411765	1
Uterine Prolapse/*epidemiology/etiology	0.011812069	5	2.03741497	0	0.49081803	1
CVP: Centralidad de valor propio. CG: Centralidad de Grado. PCMC: Promedio de caminos más cortos. CI: Centralidad de intermediación. CC: Centralidad de cercanía. CA: Coeficiente de agrupamiento.						

Cuadro 4. Características topológicas de las subredes correspondientes a la partición en comunidades.

Red	Nodos	Aristas	Coefficiente de agrupamiento	Longitud de camino	Centralidad	Número promedio de vecinos	Densidad de red
Edad mediana	62	552	0.881	1.723	0.630	17.806	0.292
Mujer	19	67	0.848	1.620	0.618	7.053	0.392
Factores de riesgo	18	153	1.0	1.0	0.0	17	1.0