




Análisis temático del discurso alimentario a partir de grafos de coocurrencias

Antonio Rifón Sánchez



Universidade de Vigo 

Este artículo se publica con posterioridad al fallecimiento del profesor Antonio Rifón Sánchez, responsable del análisis y la discusión en la que deja entrever su genialidad como lingüista y su forma de ser divertida y cercana. Era un investigador brillante y un compañero extraordinario.

Susana Rodríguez Barcia

Universidade de Vigo  

Ana Varela Suárez

Universidade de Vigo  

<https://dx.doi.org/10.5209/clac.93486>

Recibido: 1 de febrero de 2021 • Aceptado: 31 de enero de 2023

ES Resumen. Este artículo se centra en el análisis temático del discurso alimentario de la divulgación científica en los medios de comunicación generalistas. Sus objetivos son: (i) establecer los temas del discurso de la alimentación; (ii) determinar el grado de importancia y centralidad de cada uno; (iii) analizar las relaciones que se establecen entre ellos; y (iv) valorar los enfoques ideológicos que se ha dado a estos temas. Se parte del análisis de un corpus de 144 artículos sobre alimentación de los 12 periódicos generalistas con mayor alcance; de este corpus se ha confeccionado un grafo de coocurrencia de palabras no dirigido y pesado a partir del cual se han calculado el grado de intermediación, el grado del nodo y la modularidad. Estas medidas nos han permitido identificar comunidades que constituyen los distintos temas y subtemas. Se ha concluido que el enfoque fundamental de la alimentación es la salud, que los alimentos ya no son simplemente comida, sino nutrientes y que los colectivos que centran esta visión son las mujeres, los niños y las niñas.

Palabras clave: discurso alimentario, análisis temático, ideología, teoría de redes.

ENG Thematic analysis of food discourse from co-occurrence graphs

Abstract. This paper focuses on the thematic analysis of food discourse on mass media when they conduct scientific dissemination. It aims to (i) establish the topics of food discourse; (ii) determine the degree of importance and centrality of topic, (iii) analyze the relationships established among them and (iv) assess the ideological approaches given to these issues. For this purpose, a corpus with 144 news on nutrition published in the most-read Spanish general media has been compiled. From this corpus, an undirected and weighted word co-occurrence graph has been elaborated and its betweenness, degree and modularity have been calculated. These measures have allowed us to identify communities that constitute the different topics and subtopics. It has been concluded that the fundamental focus of food is health, that food is no longer perceived as just food, but as nutrients, and that the main targeted groups are children and women.

Keywords: food discourse, thematic analysis, ideology, network theory.

Sumario. 1. Introducción. 2. Antecedentes. 3. Materiales y métodos. 3.1. Diseño del corpus. 3.2. Método. 4. Resultados. 4.1. Medidas globales del grafo. 4.2. Las comunidades temáticas generales. 4.3. La comunidad alimentación. 4.4. La comunidad 2-Comer. 4.5. Las otras comunidades. 4.6. Discusión general. 5. Conclusiones. Referencias bibliográficas.

Cómo citar: Rifón Sánchez, A.; Rodríguez Barcia, S. y Varela Suárez, A. (2024). Análisis temático del discurso alimentario a partir de grafos de coocurrencias, *Círculo de Lingüística Aplicada a la Comunicación* 97 (2024), 271-288. <https://dx.doi.org/10.5209/clac.93486>

1. Introducción

La ciencia y su investigación, a pesar de su búsqueda de objetividad, claridad y precisión, está cargada de teoría y no es, tampoco, ajena a preferencias subjetivas, influencias sociales y posturas ideológicas, “perhaps we cannot sensibly draw scientific inferences without a host of background assumptions, which may include assumptions about values” (Reiss y Sprenger, 2017).

Si la propia investigación, a pesar de sus esfuerzos por lograr la objetividad, no está libre de elementos ideológicos, mucho menos lo estará su divulgación. En ella se reelaboran, se textualizan y se redennominan los discursos especializados, a través de la combinación de elementos del discurso científico y del discurso general para hacer accesible al público no especializado los conocimientos científicos (Cassany et al., 2000).

Un mismo discurso científico puede ser reelaborado de formas muy diferentes y desde distintas perspectivas. La elección de lo que se elimina y de lo que se añade puede dar una orientación de los intereses científicos de quien divulga, pero también de otros intereses, en principio, espurios en la ciencia, como los económicos, los ideológicos, etc.

En la transmisión de la ciencia existe un desequilibrio de conocimientos entre las personas que la escriben, los especialistas o semiespecialistas, y las personas legas o con conocimientos limitados que la reciben (Gouveia y Alexandre, 2009). Este desequilibrio no solo es aprovechado para la transmisión de conocimiento, sino que a través de él se transmiten también ideología, valores e intereses ajenos a la ciencia.

Así, la condición de desventaja por parte de las personas receptoras frente a las transmisoras, hará que las primeras asuman, en gran medida, la verdad de los contenidos como una condición intrínseca al conocimiento científico, y que se subjetive esa realidad asumiéndola como propia.

La ciencia abarca muchos ámbitos y la ideología se mostrará en distinto grado según la temática que se reelabore: cuanto más expuesto esté el ámbito científico a intereses ajenos a la ciencia, sus discursos se mostrarán más ideologizados. Los discursos científicos que tienen como tema principal la alimentación en cualquiera de sus dimensiones (social, nutricional, ética, ecológica, cultural, etc.) tienen la capacidad de influir directamente sobre la sociedad, bien en aspectos ideológicos sin implicación aparentemente práctica, bien en dinámicas concretas de consumo; esto supone que el ámbito alimentario sea uno de los que más expuestos esté a la presión de intereses ideológicos o comerciales.

Esta realidad es la que propicia y sustenta las investigaciones en el análisis del discurso alimentario, pues resulta imperativo que desde estas investigaciones se ayude a comprender e interpretar el tipo de mensajes que la sociedad está recibiendo sobre este tema y, muy especialmente, los que recibe a través de los textos de la divulgación científica de nivel más bajo, presente en los medios de consumo masivo.

La tensión entre la objetividad pretendida y la subjetividad inherente a los textos alimentarios hace necesario profundizar en el estudio de la presencia de ideología en los discursos mediáticos de la alimentación como manifestaciones que, por su incidencia social en las prácticas de consumo y en los hábitos de vida (Cook, 2010), constituyen un objeto de estudio fundamental para la lingüística.

Es, en este marco, en el que se encuadra este trabajo cuyo objetivo principal es identificar los temas presentes en los medios escritos generalistas con mayor difusión en España; la identificación de los temas nos permitirá establecer el enfoque general con el que estos medios tratan el campo de la alimentación.

A partir de aquí podemos concretar cuatro objetivos secundarios: en primer lugar, establecer los temas centrales de la divulgación científica sobre alimentación; en segundo lugar, determinar el grado de centralidad de estos temas para sopesar su importancia; en tercer lugar, analizar las relaciones que se establecen entre ellos; y, en cuarto lugar, hacer un análisis preliminar, a partir de los temas centrales, de los enfoques ideológicos que priman en este tipo de discurso.

Para alcanzar estos cuatro objetivos de la forma menos subjetiva posible, en este trabajo se ha partido de un tratamiento cuantitativo de los datos y, a continuación, se han analizado estos resultados desde una perspectiva cualitativa. El método cuantitativo elegido es el de grafos de coocurrencias puesto que, por un lado, permite mantener, por lo menos en la primera fase del trabajo, la objetividad necesaria para, después, hacer una interpretación más fundamentada; por otro lado, permite, a través de las métricas de la teoría de grafos, establecer las relaciones temáticas existentes en los textos; y, por último, provee de un método transdisciplinar ya comprobado en otras disciplinas.

Para desarrollar nuestro estudio, trataremos, en primer lugar, los antecedentes de la materia; en segundo lugar, explicitaremos detalladamente el corpus de textos analizado y la metodología empleada; en tercer lugar, expondremos los resultados de los grafos establecidos haciendo una primera discusión de los temas principales, para, una vez establecidos todos los resultados, hacer una discusión crítica de ellos.

El trabajo lo hemos organizado en cuatro apartados: el primero en el que se realiza un balance de los antecedentes del tema; el segundo, en el que se detallan los materiales analizados y el método de análisis; el tercero, en el que se dan los resultados y se discuten, primero de forma individualizada para, después, acabar con una discusión general en el que se hace una valoración de los resultados en su conjunto; y, por último, el cuarto, unas breves conclusiones para mostrar que se han alcanzado los objetivos y cuáles han sido las conclusiones principales.

2. Antecedentes

La mayoría de los análisis de los discursos alimentarios surgen en la primera década del siglo XXI, en la que la globalización es un fenómeno tan imparable que ha supuesto una modificación real en los hábitos de consumo y, en general, en la forma de ver y hablar de la comida; una muestra son los trabajos del lingüista Guy Cook publicados entre 2004 y 2010.

Estos estudios se han hecho desde diferentes disciplinas y combinan aspectos culturales, ideológicos y discursivos. Para ordenar un poco el conjunto, sin que sirva como una clasificación excluyente, hay trabajos que analizan los discursos desde la perspectiva de la seguridad alimentaria, la proyección de modelos dominantes o la promoción de prácticas dietéticas o las estrategias lingüísticas empleadas.

Jarosz (2011) o MacMillan y Dowler (2012, p. 201) abordan sus trabajos desde el análisis de la ideología que esconden los discursos de política de seguridad alimentaria. Por ejemplo, los últimos concluyen que el lenguaje y la formulación del discurso sobre políticas de seguridad en la alimentación revelan prioridades y objetivos por parte de los grupos de poder.

Sobre la proyección de modelos dominantes, podemos encontrar el de Lupton (2000) que parte de la noción de riesgo, un mecanismo habitual de control de la sociedad o el de (Annandale y Clark, 2000) que estudian la relación entre los discursos de la salud y las nociones de género y clase social.

Otros modelos hegemónicos proyectados por los discursos de la ciencia han llevado a cánones físicos y estilos de vida estandarizados que han sido analizados en diversos trabajos del ámbito de la comunicación y la publicidad (Díaz Rojo y Morant i Marco, 2007). Como advierten Gracia Arnaiz y Comelles (2007), desde la segunda mitad del siglo XX se ha producido un notable cambio cultural en las representaciones del cuerpo y de sus cuidados, especialmente en lo que se refiere a la mujer, aunque no de forma exclusiva. Es importante mencionar en este punto la relevancia que están adquiriendo los discursos alimentarios presentes en redes sociales, en especial en Instagram, en relación con la consolidación de estereotipos físicos y de modelos saludables. La investigación actual se está centrando en estos textos por su influencia en colectivos de jóvenes y grupos sensibles a la información nutricional. Las estrategias divulgativas cambian y se especializan en esta red a través de la multitud de perfiles de alimentación y nutrición, que en el marco español han sido analizados recientemente por Marauri Castillo et ál. (2020).

Gracia Arnaiz (2007) ya aborda el tema de la construcción social de la normalidad dietética, y aclara uno de los aspectos más polémicos: la promoción de unas prácticas y rutinas de alimentación frente a otras y la naturalización cultural de esas prácticas, que son asumidas como la única opción saludable posible. Los trabajos sobre la construcción de la alimentación saludable en la publicidad son muy numerosos, y en el ámbito hispanohablante destaca, por ejemplo, la aportación de Díaz Rojo, Morant y Westall (2005). En cuanto a la naturalización ideológica de los discursos alimentarios dominantes, Ferreira da Silva (2013: 250) comenta el sometimiento a la visión hegemónica de la ciencia que domina en las sociedades actuales en lo relativo a las normas de alimentación variada, pretendidamente saludable, que tienden además hacia la homogeneización de prácticas.

Esto lleva a una de las grandes temáticas en las investigaciones sobre el discurso de la alimentación, las estrategias lingüísticas que envuelven los géneros discursivos en los que se trata de la alimentación ecológica, también denominada biológica (Nowacek y Nowacek, 2008). La polarización que actualmente se observa en los discursos sobre la alimentación ecológica posee un interés especial en España, ya que en la divulgación científica se está asistiendo al choque, con repercusión mediática, entre críticas feroces al respecto de los beneficios que reporta el consumo ecológico (Mulet, 2011) y textos que avalan las bondades de una alimentación con gran presencia de los productos ecológicos. Ya se mencionó que la argumentación discrepante entre defensa y detracción, tanto en lo que respecta al consumo ecológico como, por mencionar otras problemáticas, al carácter saludable e impacto medioambiental de los alimentos transgénicos, por ejemplo, ha sido ampliamente tratada en algunas investigaciones como la de Cook (2004).

Para terminar el panorama general de investigaciones realizadas sobre ideología y discurso alimentario, hay que señalar que en el ámbito hispanohablante son más los trabajos que se han realizado en países americanos que en España. En España, el estudio del discurso alimentario es todavía una línea de investigación reciente. Si bien es cierto que hay multitud de investigaciones realizadas desde la perspectiva antropológica (de Benito Morán, 2016), de la comunicación y la publicidad (Díaz Rojo et al., 2005; Díaz Rojo y Morant i Marco, 2007), no se observa el mismo desarrollo de la investigación en el ámbito de la lingüística, en el que, a pesar de algunos estudios plenamente lingüísticos como, por ejemplo, el de (Cassany y Martí, 1998), faltan estudios globales que den cuenta del panorama temático conjunto de los discursos alimentarios, especialmente de los discursos de divulgación publicados en los medios, como es el caso de Varela, Rodríguez Barcia y Rifón (2020) y del presente estudio. López Torres y Contreras Hernández (2020) han identificado, en una investigación sobre la percepción de la alimentación saludable por parte de un grupo de mujeres, la presencia masiva de discursos que implican una *nutricionalización* de la alimentación, y también detectaron procesos de asimilación y resignificación que determinan la percepción de “lo saludable” y condicionan las prácticas culinarias.

El método empleado para el análisis se puede relacionar con la extracción de temas (*topics*) y palabras clave (*keywords*) en el procesamiento del lenguaje natural y, en este ámbito, hay ya numerosos trabajos que emplean de alguna forma el análisis de grafos para dicha extracción (Mihalcea y Radev, 2011; Lahiri, 2013; Zhou et al., 2013; Bougouin et al., 2013; Boudin, 2018; Garg y Kumar, 2018), pero también con otras áreas como la calidad textual y el análisis estilístico por medio de grafos (Antiqueira et al., 2005, 2007; Amancio, 2015). De los primeros tomamos la idea de que los grafos son una herramienta adecuada para analizar la temática de los textos, de los segundos que sirve también para el análisis de la organización textual. Un punto de partida de nuestra aplicación de los grafos al análisis textual es Paranyushkin (2011) cuya intención es rastrear las huellas, “pathways”, que generan significado en la construcción del texto a través del establecimiento de grupos de conceptos por su grado de relación y la identificación de aquellos que más peso tienen en la construcción del significado textual.

3. Materiales y método

En este apartado expondremos, en primer lugar, cómo fue elaborado el corpus y sus características finales, las cuales determinan el alcance y aplicación de la investigación, y, en segundo lugar, explicitamos la forma

en que han sido creados los distintos grafos para que se comprendan mejor los resultados obtenidos y su interpretación.

3.1. Diseño del corpus

El corpus creado para este trabajo está constituido por discurso alimentario presente en la versión electrónica de medios periodísticos generalistas. Para su recopilación, se decidió limitar la búsqueda a aquellos diarios que, según el Resumen del Estudio General de Medios (octubre 2017-mayo 2018) (AIMC, 2018), cuentan con más de 200.000 lectores/as por día; de estos, se eliminaron aquellos cuya temática era exclusivamente deportiva, pues el objetivo era limitarse solo a diarios generalistas.

Todos los artículos que se han recogido han sido publicados entre el 22 de mayo y el 7 de julio de 2018. El sistema de búsqueda ha consistido en la introducción del término “alimentación” en el motor de búsqueda propio de los diarios y, a continuación, se ha hecho una selección entre los resultados en los que la alimentación era el tema central desechando aquellos que, si bien trataban la alimentación, este era un tema tangencial. Esta selección produjo un corpus de 144 artículos de 12 medios diferentes (Tabla 1).

Tabla 1. Información sobre el corpus textual recopilado con el medio consultado, el número de lectores/as en miles y el número de noticias cuyo tema central era la alimentación

Medio	Lectores/as EGM/día (000)	Noticias recopiladas
<i>El País</i>	1128	19
<i>El Mundo</i>	701	6
<i>La Vanguardia</i>	612	11
<i>La Voz de Galicia</i>	531	14
<i>Abc</i>	429	30
<i>El Periódico</i>	363	1
<i>El Correo</i>	358	5
<i>La Nueva España</i>	286	10
<i>Faro de Vigo</i>	235	6
<i>Heraldo de Aragón</i>	235	27
<i>La Razón</i>	215	13
<i>Levante</i>	210	2
Total	-	144

La disparidad en el número de textos recopilados en cada medio no es, para nuestra investigación, importante, ya que para nuestros objetivos actuales no tiene importancia en qué medio concreto fue publicada la noticia.

Las medidas globales de estos 144 textos pueden verse en la Tabla 2:

Tabla 2. Datos generales del corpus: número de oraciones que contiene, número de párrafos, número de palabras, número de palabras distintas y lemas, entendidos como los distintos temas de las palabras, es decir, las palabras desprovistas de morfemas flexivos.

Medidas	Oraciones	Párrafos	Palabras	Palabras distintas	Lemas
Corpus	4325	1901	108495	11364	7070
Media	31,06	14,20	829,07	315,27	271,08
Mediana	24	12	688,5	284,5	250,5
Curtosis	13,97	12,12	14,99	9,28	7,76
Asimetría	3,28	3,01	3,08	2,25	1,96

En la Tabla 2 se muestra que el corpus es lo suficientemente amplio para poder hacer una aproximación a la divulgación en los medios masivos generalistas. Los textos recopilados no son ni excesivamente grandes, ni excesivamente cortos; una media de 14,2 párrafos representa bien el género que estamos tratando.

Por otro lado, los coeficientes de curtosis y asimetría pequeños indican que los datos se agrupan mayoritariamente hacia la media; por ejemplo, en el caso del número de párrafos tenemos una distribución ligeramente leptocúrtica, es decir, ligeramente apuntada, por lo que la mayoría de los datos se agrupan hacia la media con una ligera asimetría a la derecha; esta asimetría positiva indica que el agrupamiento se hace hacia valores menores que la media. Como en ambos casos los valores son pequeños, podemos concluir

que los textos son equilibrados en cuanto a sus valores globales: no tenemos muchos textos muy largos y unos pocos muy cortos, ni muchos muy cortos y unos pocos muy largos.

Este equilibrio es importante para la investigación pues un desequilibrio grande, debido a la naturaleza del método, podría darnos resultados engañosos, ya que, si el corpus estuviese compuesto por unos pocos textos muy largos y muchos cortos, los primeros podrían ocultar las características de los segundos y, si la situación fuera la inversa, las relaciones temáticas podrían parecer mucho más segmentadas de lo que realmente son.

Lo dicho para el número de párrafos sirve para las otras medidas, por lo que no profundizaremos más en el análisis de los datos. Pasemos ahora a analizar cómo se procesaron los textos para su posterior análisis.

3.2. Método

El objetivo de todos los pasos que explicitaremos será obtener un grafo no dirigido ponderado de coocurrencias textuales.

- Lematización:** una vez recopilados los textos, estos se han lematizado a través de *Freeling* (Carreras et al., 2004; Padró y Stanilovsky, 2012) lo cual produce un documento en el que tenemos la palabra gramatical, su categoría y el lema desprovisto de valores gramaticales. Estos datos han sido exportados a una base de datos sql. En esta base se recogen los tres datos de la lematización (palabra, categoría y lema), el texto y el orden en que aparecen y se les asigna el número de oración y párrafo en que aparecen. Una vez recogidos estos datos, se ha procedido a la corrección manual de algunos pocos errores de lematización, fundamentalmente unificando nombres propios de personas y palabras extranjeras y técnicas.
- Generación del texto:** en nuestro método no nos interesan todas las categorías gramaticales; nuestro análisis se centra en la coaparición de sustantivos, adjetivos y verbos. Para ello se genera una nueva tabla sql en el que solo aparecen estas categorías con la oración y párrafo de aparición, eso sí, manteniendo el orden de aparición para poder establecer después las relaciones.
- Generación del grafo:** para generar el grafo, se genera, en primer lugar, una tabla sql de nodos y después una con las aristas.

Para generar los nodos solo hay que extraer todos los lemas diferentes de la tabla anterior, es decir, todos los sustantivos, adjetivos y verbos diferentes. A cada uno de estos lemas se le ha añadido información sobre los textos y párrafos en que aparecen.

Para generar la lista de aristas hay que tener en cuenta, en primer lugar, qué relaciones queremos reflejar. Hemos optado por considerar las relaciones de coaparición restringidas al párrafo; es decir, no se establecen relaciones entre la última palabra de un párrafo y la primera del siguiente, lo cual implica que la última palabra de un texto y la primera del siguiente tampoco se relacionan.

En segundo lugar, hemos considerado que las relaciones de contigüidad tienen mayor peso que las de coaparición no contigua, a las primeras les hemos dado un valor de 2 y a las segundas de 1.

En tercer lugar, se ha considerado una ventana de 5-gram, es decir, la palabra considerada establece relaciones con las dos anteriores y con las dos siguientes. Por ejemplo, en la frase “(el) libro rojo *está* (en la) mesa (de) Juan”, siendo *está* la palabra a partir de la que queremos establecer las relaciones, según nuestro análisis se relaciona con *rojo* y *mesa* con un valor de 2 y con *libro* y *Juan* con un valor de 1.

Establecido esto, solo queda recorrer los textos estableciendo las relaciones y su peso y relacionarlas con los nodos, de manera que los nodos estarán relacionados entre sí independientemente del texto en que aparezcan y tendrán mayor grado de relación dependiendo del número de veces que coaparezcan y del grado de cercanía de esas coapariciones.

De esta manera obtenemos lo que deseamos: un grafo no dirigido ponderado o pesado de coocurrencias textuales con ventana de 5-gram en el que el corpus de 144 textos es analizado como una única unidad textual. Construido el grafo, se exporta al programa de análisis de grafos Gephi (Bastian et al., 2009) para su análisis. Si se quiere conocer con más detalle el método, se puede ver una explicación más extensa en Rifón (2020).

4. Resultados

En primer lugar, analizaremos el grafo de forma global para obtener una visión de conjunto, para lo cual nos centraremos en las medidas de la totalidad del grafo y del gran componente conexo que lo conforma. A continuación, analizaremos la modularización de este componente para delimitar sus comunidades (o comunidades temáticas). Por último, procederemos al análisis de estas últimas.

4.1. Medidas globales del grafo

En la Tabla 3, se encuentran detalladas las medidas del grafo completo; lo primero que llama la atención es que esté compuesto por 3 componentes conexos, es decir hay tres grupos de nodos que no están conectados entre sí. Por la manera en que se ha hecho el grafo, esto no debería ocurrir, a no ser que haya párrafos cuyos sustantivos, verbos y adjetivos no se repitan en ningún otro párrafo. Uno de los tres componentes es el título de un apartado, “halitosis”, palabra que no se vuelve a emplear en ningún texto, y otro es un elemento de

una enumeración cuyas palabras no vuelven a aparecer: “ciruela claudia (calle) pano (calle) ruata”. Nuestro análisis dejará a un lado estos dos elementos y se centrará en el gran componente conexo que contiene el 99,92% de los nodos y en sus medidas (Tabla 3). Explicamos a continuación la terminología de la tabla. Los *nodos*, también denominados *vértices*, remiten en la teoría de grafos a la unidad fundamental que conforma los grafos o imágenes. Las *aristas* son las líneas que unen dos vértices de un grafo, y equivalen a una relación. Pueden ser dirigidas o no dirigidas, dependiendo de si se les asigna o no un sentido. El *diámetro* equivale a la mayor distancia entre dos puntos o vértices del grafo. La *densidad* remite a la proporción de aristas que posee el grafo. Un grafo será más denso cuantas más aristas presente, y será disperso cuando presente un número bajo de aristas. La *modularidad* mide la estructura de las redes o grafos, en concreto, la fuerza de la división de una red o grafo en comunidades (también denominadas clústeres o módulos).

Tabla 3. Principales medidas del grafo completo y de su gran componente conexo (99,92% de los nodos).

	Grafo completo	Gran componente conexo
Nodos	6039	6034
Aristas	61027	61022
Grado medio	20,211	20,226
Grado medio con pesos	70,715	70,769
Diámetro	9	9
Densidad	0,003	0,003
Modularidad	0,235	0,244
Componentes conexos	3	1
Coefficiente de clustering	0,402	0,402
Longitud media de camino	3,036	3,036

De este gran componente conexo, hay dos medidas que merecen una discusión: la densidad y la modularidad. Para ello, se ha de tener presente que los textos tienen una temática común: la alimentación, por lo que el léxico está restringido a ese campo y, por tanto, tendrán palabras comunes. Igualmente, se debe considerar que el campo temático es muy general, por lo que cabe esperar que haya diferencias léxicas entre algunos subtemas, aunque también estén unidos por un cierto número de léxico común.

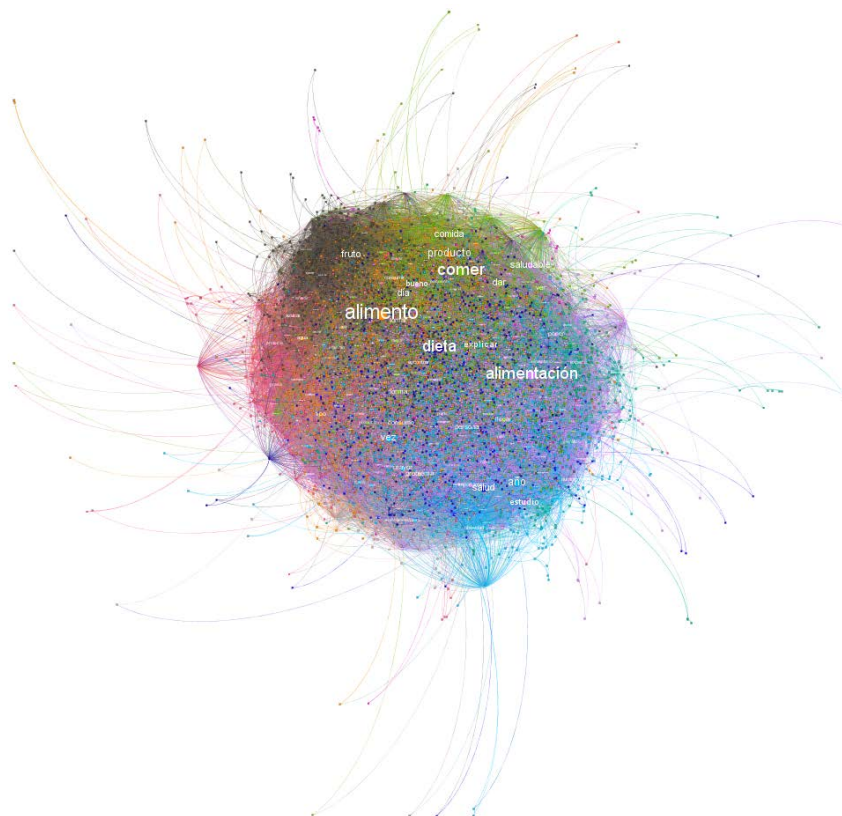


Figura 1. Grafo del componente conexo. Los colores representan comunidades y el tamaño de la etiqueta de los nodos está escalado por su grado de intermediación (Distribución ForceAtlas-2).

La tensión entre unidad temática general y amplitud temática en los subtemas, se puede ver en la representación del grafo (Figura 1). Existe un núcleo central compacto cohesionado por algunas palabras comunes (alimento, dieta, alimentación, comer, comida, fruto, estudio...), no todas localizadas en el centro del núcleo, sino distribuidas en diferentes lugares del mismo, así como pequeños brazos que parten de ese núcleo hacia el exterior. Las palabras que se distribuyen en diferentes lugares del núcleo central cumplen dos funciones: por un lado, cohesionar los textos en torno a un tema general; por otro, distinguir los subtemas de ese tema central. Los brazos representan subtemas o subtemas de subtemas que aparecen en textos particulares.

Esta situación explica su baja densidad que indica que, en su interior, se mantienen pocas conexiones entre palabras frente a todas las posibles; si bien una densidad relativamente baja es normal en los grafos de coapariciones de palabras, en este caso un valor tan bajo merece algún comentario. Durante la recopilación de los 144 textos del corpus se partió de la palabra “alimentación”, con el objetivo de obtener un amplio abanico de temas y enfoques. Esto provoca que, como ya hemos mencionado, exista una gran variedad de subtemas. Por lo tanto, la repetición léxica entre ellos es menor que en corpus más homogéneos temáticamente. Además, como se reduce la repetición léxica también lo hace el número de relaciones establecidas y, con ello, la densidad del grafo.

También presenta una baja modularidad, ya que este dato nos indica lo fácil o difícil que resulta distinguir comunidades de palabras que mantienen entre ellas relaciones más densas que la media del grafo y son “groups of vertices which probably share common properties and/or play similar roles within the graph” (Fortunato, 2010, p. 3). En estos textos tenemos un grupo de palabras que aparecen en todos los subtemas y que, por tanto, hacen más difícil distinguir comunidades; si los textos tuvieran temáticas muy diferenciadas, alimentación y volcanes, por ejemplo, estas palabras comunes generales no existirían y el grado de modularidad aumentaría.

En la Figura 1 las diferentes comunidades se representan en diferentes colores y, si estuviesen claramente definidas, debería tener ciertas distancias entre ellas, pero, como las diferencias entre las comunidades no son grandes, aparecen unidas en un núcleo compacto lo que muestra que, si bien es posible distinguir comunidades, no se puede hacer de forma clara, por lo que el grado de modularidad, la capacidad para distinguir comunidades, es relativamente bajo.

Vayamos ahora al análisis de las comunidades propuestas para este gran componente. Nuestro método será, ahora, dividir para vencer; analizaremos la organización de los componentes dividiendo, algunos de ellos, en nuevos componentes o subcomponentes que faciliten el análisis temático, pero, con el fin de discernir cuáles de esas comunidades y subcomunidades forman parte del núcleo temático del corpus y cuáles de su periferia y, por tanto, pueden ser tomadas como temáticas particulares, vamos a ir reduciendo el detalle del análisis según la importancia de la comunidad.

4.2. Las comunidades temáticas generales

En la Figura 2 se recogen las 13 comunidades reconocidas por la modularización del grafo.

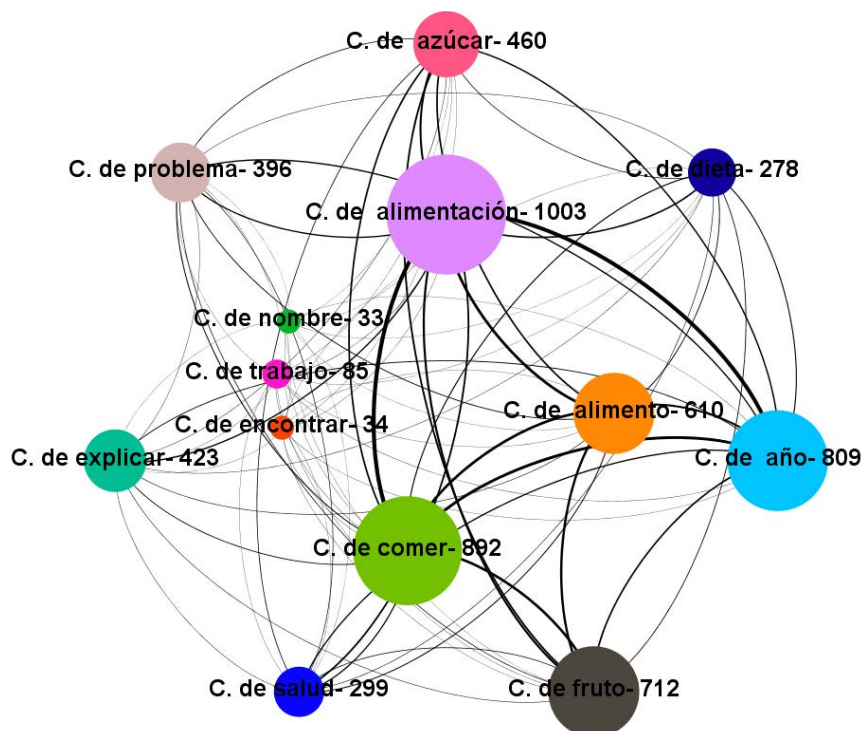


Figura 2. Grafo con las comunidades clusterizadas en el gran componente (El tamaño de los nodos representa el tamaño de la comunidad; la diferencia de grosor de la arista, el grado de relación entre comunidades y; la etiqueta, la palabra con mayor grado

La etiqueta para denominar las comunidades (Figura 2) es la palabra (desde ahora emplearemos siempre la forma lematizada en masculino singular y el infinitivo verbal) con mayor grado de intermediación de la comunidad, por lo que, en algunos casos, es temáticamente poco transparente. La intermediación (*betweenness*) (Brandes, 2001; Newman, 2010:185-193) es una medida de centralidad que mide los caminos más cortos que pasan por un nodo; los nodos con una alta intermediación unen grupos de nodos no relacionados directamente.

Volviendo a nuestro grafo, en la Figura 1 se puede observar que no todas las comunidades presentan una palabra con un alto grado de intermediación en el conjunto total y que la mayor parte de palabras con mayor grado, se concentran hacia arriba del centro del grafo; esto quiere decir que hay comunidades que concentran la mayor parte de palabras con alto grado de intermediación y comunidades que presentan pocas.

Si consideramos que la intermediación mide el grado de las palabras por su función de conectoras entre distintos grupos, podemos considerar que, cuanto más alto sea el grado de intermediación de una palabra, está representará mejor a los grupos que une. Por ejemplo, la palabra *comer* (Figura 1) unirá aquellos conjuntos de palabras entre los que *comer* sea un concepto central común, por lo que podríamos decir que *comer* representa adecuadamente a esos grupos. Por eso, ahora atenderemos a la importancia en el corpus de cada una de las 13 comunidades.

Para ello, analizaremos el grado de intermediación de las palabras de cada comunidad. En la Tabla 4, se recoge el grado de intermediación de las ocho palabras con mayor grado de intermediación del grafo, aquellas que tienen un grado mayor que 0,02; además, se recoge también el grado de intermediación de las palabras que se han tomado como etiquetas de las comunidades y que no coinciden con ninguna de las primeras ocho palabras anteriores. De ellas se da el puesto que ocupan en el grafo según su grado de intermediación.

Tabla 4. Grado de intermediación de las 8 palabras con grado mayor de 0,02 del grafo y de las palabras con mayor grado en cada comunidad indicando el puesto que ocupan en el grafo según su intermediación.

Puesto	Palabra	Intermediación	Puesto	Palabra	Intermediación
1	alimento	0,041686	11	fruto	0,019461
2	comer	0,035646	13	salud	0,019014
3	alimentación	0,034477	15	explicar	0,017609
4	dieta	0,033562	20	problema	0,014081
5	año	0,024398	27	azúcar	0,010625
6	producto	0,023889	33	encontrar	0,009584
7	comida	0,020072	127	trabajo	0,003769
8	saludable	0,020024	317	nombre	0,001331

Vemos que, en el grafo, hay ocho palabras con un valor de intermediación mayor que 0,02 y que, en estas ocho palabras, no están representadas todas las comunidades; para encontrar las palabras pertenecientes a alguna de las comunidades, hay que ir a puestos más allá del 8, por ejemplo, al puesto 317, para la comunidad *nombre*.

Esta situación nos permite ya hacer una pequeña limpieza en las comunidades que vamos a analizar; podemos considerar que las últimas tres comunidades no son generales al corpus sino temas particulares poco unidos al núcleo central; en las que, a su bajo grado de intermediación, se une que están compuestas por pocas palabras (34, 85, 33). Así que, para desbrozar un poco el camino, nos quedaremos con 10 comunidades.

Antes de pasar a analizar cada una de estas 10 comunidades, haremos una valoración temática general aprovechando el grado de intermediación; para ello, retomamos sus palabras etiqueta y les añadimos aquellas otras que tienen un grado de intermediación mayor que 0,01. Agrupamos la lista obtenida por categoría y cada categoría la ordenamos por el grado de intermediación:

Nombres: alimento, alimentación, dieta, año, producto, comida, vez, fruto, día, salud, estudio, persona, forma, problema, consumo, tipo, azúcar, experto, obesidad, nutrición, agua.

Verbos: comer, dar, explicar, ver, tomar, poner, llegar, consumir.

Adjetivos: saludable, bueno, mayor, importante.

Este conjunto de palabras da una visión temática global del corpus. El corpus se ocupa del tema *alimentación-alimento* enfoca este hacia la *nutrición* tratando la *dieta* y el *consumo* del alimento como *producto* o *fruto* para cuidar *problemas* que afectan a la *salud* como la *obesidad* y se apoya en *estudios de expertos*. Todos estos temas se enfocan como elementos *importantes* pues son *saludables* y *buenos*.

Ahora bien, esto es muy general, debemos analizar con más detalle cada una de las comunidades para establecer los grupos temáticos y el tratamiento dado a cada uno. Analicemos, entonces, los datos generales y algunos aspectos gráficos de las 10 comunidades (Tablas 5 y 6):

Tabla 5. Medidas globales de las primeras cinco comunidades del corpus.

Medidas Comunidades	1-Alimentación	2-Comer	3- Año	4-Fruto	5-Alimento
Nodos	1003	892	809	712	610
Aristas	4839	3528	2956	2817	2021
Grado medio	9,649	7,910	7,308	7,913	6,626
Grado medio con pesos	34,883	27,231	27,718	31,045	24,816
Diámetro	6	8	8	9	9
Densidad	0,010	0,009	0,009	0,011	0,011
Modularidad	0,358	0,358	0,429	0,424	0,436
Componentes conexos	1	1	1	1	1
Coefficiente de clustering	0,473	0,485	0,504	0,525	0,534
Longitud media de camino	3,071	3,204	3,364	3,282	3,274

Tabla 6. Medidas globales de las cinco últimas comunidades del corpus.

Medidas Comunidades	6-Azúcar	7-Explicar	8-Problema	9-Salud	10-Dieta
Nodos	460	423	396	299	278
Aristas	1562	855	944	712	651
Grado medio	6,791	4,043	4,768	4,763	4,683
Grado medio con pesos	27,265	13,835	17,707	18,836	18,072
Diámetro	8	11	9	8	8
Densidad	0,015	0,010	0,012	0,016	0,017
Modularidad	0,475	0,701	0,560	0,559	0,545
Componentes conexos	1	1	1	1	1
Coefficiente de clustering	0,546	0,659	0,549	0,578	0,554
Longitud media de camino	3,303	3,848	3,671	3,541	3,306

De estas medidas, queremos destacar el aumento de la modularidad y del coeficiente de *clustering* con respecto al grafo completo. El aumento en la modularidad indica que según avanzamos en las comunidades, estas son más fáciles de modularizar, es decir, los temas o subtemas que tratan están más diferenciados. El aumento del coeficiente de *clustering* nos indica que las palabras forman grupos o subgrupos más compactos, ya que este coeficiente mide la media en la que los nodos vecinos de un nodo son, a la vez, vecinos entre sí. Así que tenemos, por un lado, temáticas más diferencias y, por otro, más cohesionadas internamente, esto nos permite distinguir el grado de generalidad o particularidad de las comunidades hasta ahora reconocidas.

Dejemos un momento los datos y vayamos a ver las repercusiones en la representación gráfica; para ello hemos tomado cuatro comunidades, la de valores menores de modularidad *-1-Alimentación* (Figura 3)-, la de valores intermedios *-5-Alimento* (Figura 4)-, la de mayores valores *-10-Explicar* (Figura 5)- y la anterior a esta *-9-Salud* (Figura 6).

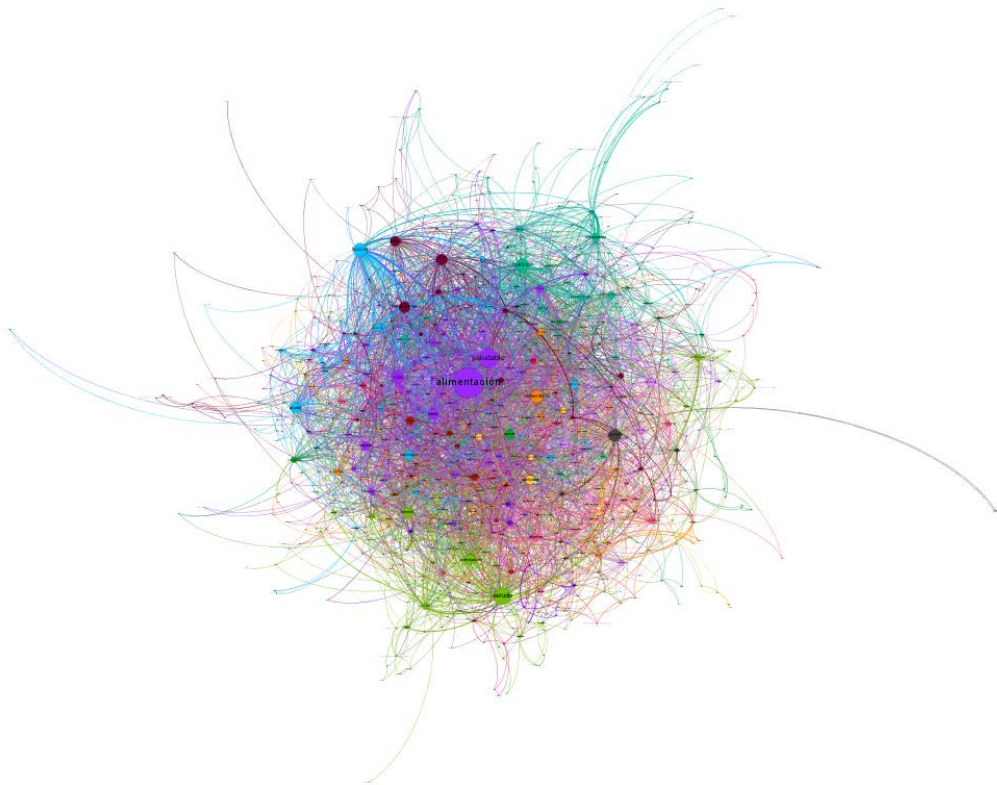


Figura 3. Grafo de la comunidad Alimentación.

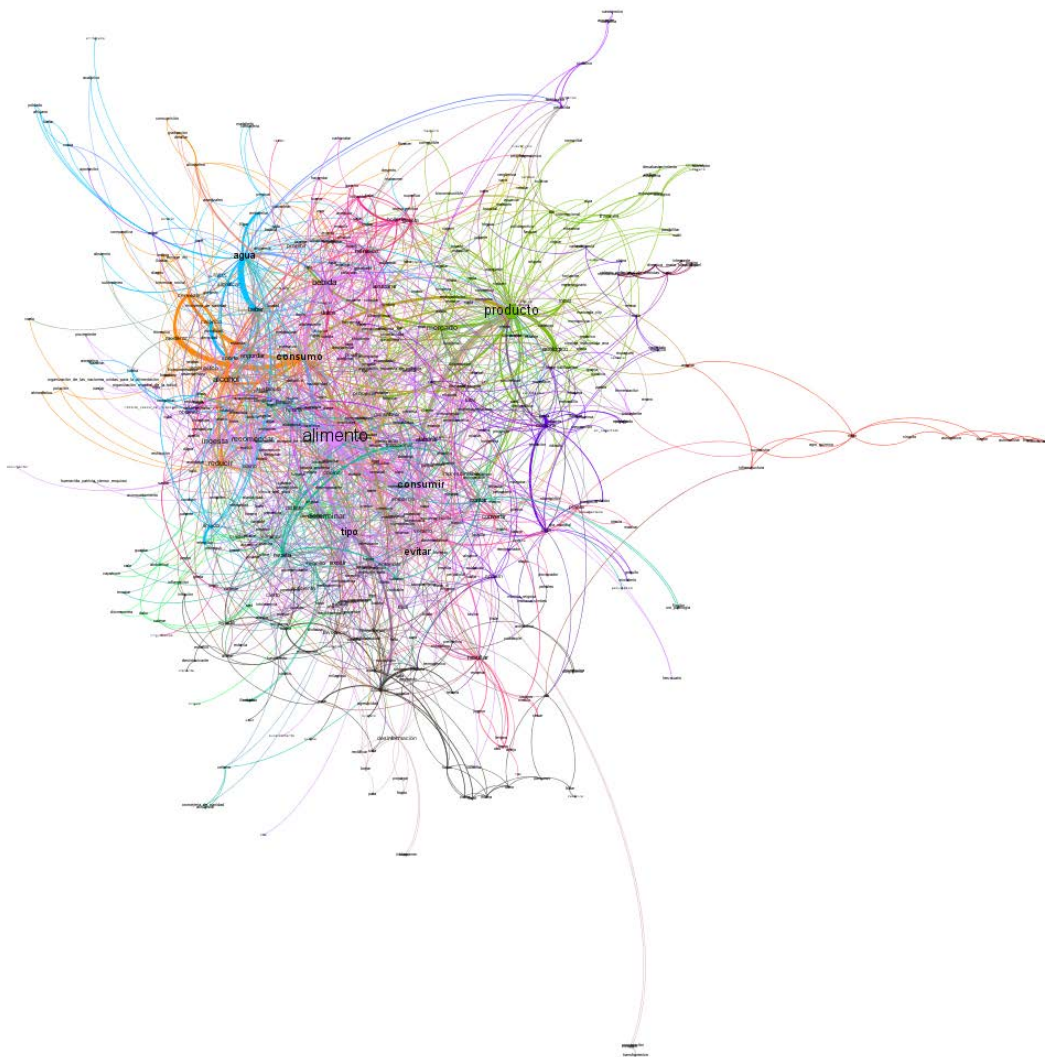


Figura 4. Grafo de la comunidad Alimento.

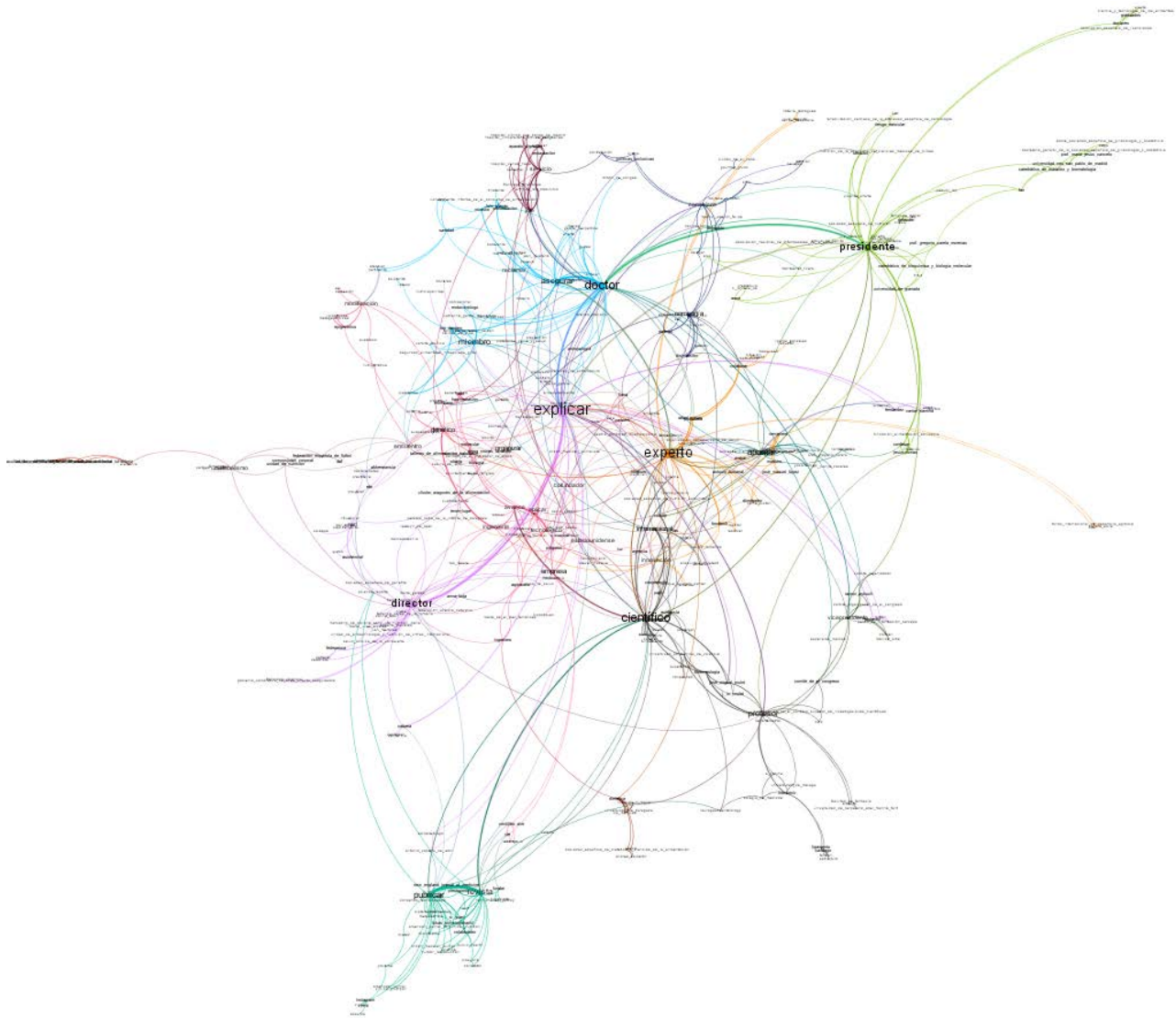


Figura 5. Grafo de la comunidad Explicar.

En estas cuatro figuras se muestra claramente que según aumenta la modularización y el grado de clusterización, las comunidades son más claras y están más separadas. En el primer grafo, vemos un núcleo central denso con algunos brazos que salen de él, pero pocos; un grafo muy parecido al del grafo completo (vid. Figura 1); en el segundo grafo, el centro comienza a dispersarse, se pueden diferenciar por lo menos dos zonas, y los brazos aumentan tanto en número como en tamaño y en distancia del centro; en el tercero, es ya difícil reconocer un centro claro, si bien parece que en torno a *salud* hay cierta aglomeración, esta no está en el centro del grafo y podemos ya distinguir otros centros; en el cuarto, ya no hay un único núcleo y cada uno de esos núcleos tiene sus propios brazos.

Esta, podemos llamarla, descentralización o desnuclearización indica que según avanzamos en las comunidades atendiendo a estas medidas, nos movemos de una temática general a una más local o particular; por eso, vamos a optar, al estudiar las comunidades por variar también nuestro grado de detalle. De la primera comunidad, *Alimentación*, analizaremos sus datos globales y los datos de las subcomunidades que la componen; de la segunda comunidad, *Comer*, analizaremos los datos globales y algunas de sus subcomunidades; del resto de comunidades solo analizaremos cada comunidad como un todo, sin atender a sus diferentes subcomunidades pues, hacerlo, implicaría dar entrada en el análisis a temas locales o particulares de los textos y esto queda fuera de nuestros objetivos.

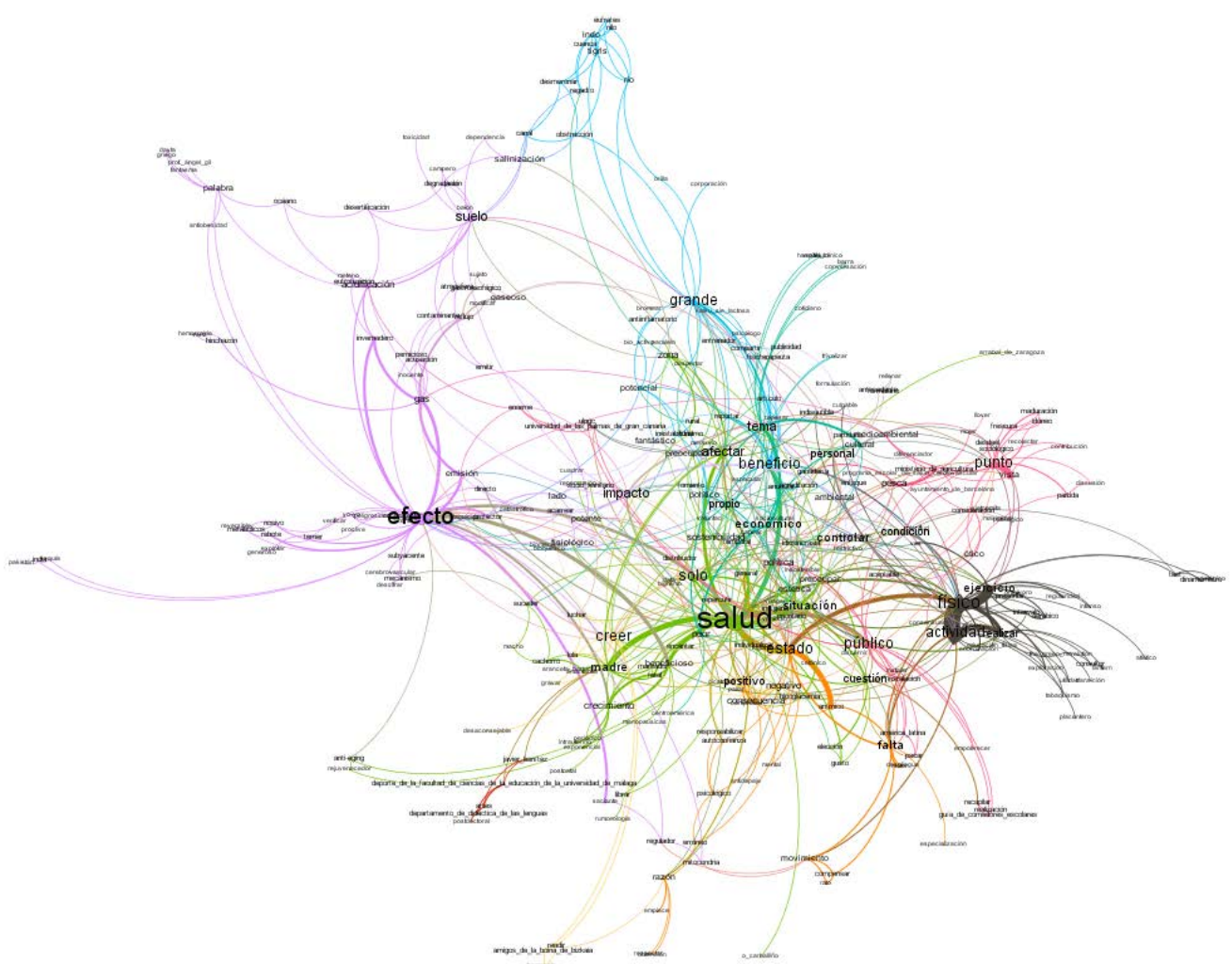


Figura 6. Grafo de la comunidad Salud.

4.3. La comunidad alimentación

Como ya se ha visto en las medidas globales (Tabla 5), las subcomunidades de “alimentación” están ligeramente más diferenciadas que las del grafo completo pero menos que las del resto de comunidades y, por otro lado, es una comunidad con bastantes nodos/palabras (1003). Para su análisis procederemos a dividir y subdividir para tratar de discriminar los temas y subtemas centrales del corpus de aquellos que podemos considerar locales o particulares de los textos.

En primer lugar dividiremos la comunidad en 14 subcomunidades (Tabla 7); de estas 14 comunidades volveremos a modularizar las dos primeras con el objetivo de poder determinar con más claridad sus temas, dentro de estas subcomunidades escogeremos las palabras que tengan un grado superior a 9; del resto de subcomunidades se escogerán directamente las palabras con un grado mayor de 9, sin entrar en más detalles.

Tabla 7. Comunidades detectadas dentro de la comunidad 1-Alimentación con el tamaño (número de palabras) y el tanto por ciento que ocupan en la comunidad (En la columna Id el primer dígito indica que pertenece a la comunidad general del grafo 1-Alimentación y el segundo, el número de subcomunidad dentro de esta).

Subcomunidades de 1-Alimentación			
Id	Subcomunidad	Tamaño	%
1.1	C. de alimentación	130	12,96
1.2	C. de nutrición	125	12,46
1.3	C. de estudio	110	10,97
1.4	C. de bueno	94	9,37
1.5	C. de objetivo	88	8,77
1.6	C. de nutricional	67	6,68
1.7	C. de alimentario	63	6,28

Subcomunidades de 1-Alimentación			
Id	Subcomunidad	Tamaño	%
1.8	C. de importante	60	5,98
1.9	C. de menú	55	5,48
1.10	C. de poner	51	5,08
1.11	C. de nuevo	48	4,79
1.12	C. de necesario	45	4,49
1.13	C. de desarrollar	42	4,19
1.14	C. de destacar	25	2,49
Total		1003	100

Como se puede ver, hay comunidades, por tamaño, más grandes y se puede decir que más importantes temáticamente. Como ya se ha indicado, procederemos a modularizar las dos primeras subcomunidades *-1.1-Alimentación y 1.2-Nutrición-*; hecho esto, solo se atenderá, después solo a aquellas palabras que presentan un grado mayor que 9 (el grado aparece al lado del lema entre paréntesis). La modularización de estas dos comunidades es necesaria porque si procedemos a escoger directamente las palabras con grado mayor que 9 lo que obtenemos es una lista poco ordenada y coherente de palabras:

alimentación (65), saludable (54), vida (39), hábito (35), mejorar (20), mantener (20), importancia (15), activo (15), principio (14), fomentar (14), buscar (13), clave (13), adquirir (13), Futuro (12), alimenticio (11), estilo (11), funcionar (11), promover (10), deportivo (10), éxito (10), esencial (10).

Su modularización nos permite agrupar las palabras de esa lista y aumentar la coherencia.

También vamos a cambiar la medida a la que atenderemos; hasta ahora, hemos puesto nuestra atención en la intermediación, ahora lo haremos en el grado (número de relaciones que mantiene la palabra). Este cambio se debe a que, en este momento, comprobar las palabras que tienen más relaciones dentro de cada subcomunidad y no las que conectan en mayor medida los grupos.

Pasemos ahora a analizar nuestra primera subcomunidad, *1.1-Alimentación*, veamos ahora en qué subcomunidades se agrupan las palabras que tienen un grado mayor que 9. La lista formada está dividida en cada una de las 7 subsubcomunidades detectadas pero, como se puede ver no todas tienen la misma importancia, la subsubcomunidad *1.1.7* no tiene ninguna palabra con grado mayor que 9 (por eso queda fuera del análisis), la *1.1.5*, solo dos palabras, y *1.1.6*, solo una:

Comunidad 1.1.1: alimentación (65), saludable (54), futuro(12), promover (10).

Comunidad 1.1.2: mejorar (20), buscar (13), clave (13), deportivo (10), éxito (10).

Comunidad 1.1.3: vida (39), hábito (35), adquirir (13), alimenticio (11), estilo (11).

Comunidad 1.1.4: mantener (20), importancia (15), funcionar (11), esencial (10).

Comunidad 1.1.5: activo (15), principio (14).

Comunidad 1.1.6: fomentar (14).

Parece que, de esta manera, la lista queda reducida a 4 subcomunidades y las temáticas empiezan a cobrar más sentido. Las tres primeras subsubcomunidades nos indican los caminos temáticos que se han recorrido en esta subcomunidad. Las palabras empleadas son muy significativas, la *alimentación saludable* que se ha de *promover* hacia el *futuro* como *clave* que se ha de *buscar* para *mejorar* para *adquirir hábitos y estilos de vida* y *alimenticios*. La comunidad *1.1.4* es también relevante dentro del corpus general, tenemos *importancia* y *esencial* por un lado y *mantener* y *funcionar* por otro. Las otras dos comunidades tienen poca importancia y podemos obviarlas.

En la subcomunidad, *1.2-Nutrición*, procederemos de la misma manera, tomaremos la lista de palabras con mayor grado que 9 y para mejorar la coherencia las agruparemos por comunidades; en este caso, al limitar el grado del nodo se eliminan de nuestro foco más subsubcomunidades que consideraremos locales. Nuestra agrupación de palabras queda reducida a tres de ellas:

Comunidad 1.2.1: nutrición (30), especialista (22), endocrinología (14), paciente (10)

Comunidad 1.2.4: médico (22), área (18), profesional (18), conocimiento (16), sanitario (10)

Comunidad 1.2.6: nutricionista (32), dietista (13), facilitar (11).

Esta comunidad trata sobre los aspectos técnicos de la nutrición haciendo referencia a los especialistas que se tornan en autoridades que el paciente debe seguir, destacando su profesionalidad.

Del resto de subcomunidades de esta primera comunidad (*1-Alimentación*) no haremos modularización pues la lista de palabras con grado mayor que 9 es reducida y no es necesario agruparlas para ganar claridad:

Comunidad 1.3-Estudio: estudio (50), investigación (24), resultado (17), cambio (16), participar (12), demostrar (12), conclusión (12).

Comunidad 1.4-Bueno: ayudar (26), bueno (25), cuerpo (17), conseguir (15), señalar (12), recordar (12), perfecto (11), ejemplo (10), muscular (10).

Comunidad 1.5-Objetivo: proyecto (17), iniciativa (17), esfuerzo (15), objetivo (12), lograr (12), eduksano (11), final (10).

Comunidad 1.6-nutricional: nutricional (20), desarrollo (16), recomendación (10), específico (10)

Comunidad 1.7-Alimentario: alimentario (20), sistema (14), utilizar (13), sector (13)

Comunidad 1.8-Importante: niño (17), importante (16), parte (15), padre (15)

Comunidad 1.9-Menú: escolar (19), colegio (14), menú (13), educativo (10)

Comunidad 1.10-Poner: poner (26), valor (12)

En estas subcomunidades se puede observar los subtemas que dentro del tema general “alimentación” trata el corpus. Se apoya en los resultados, demostraciones y conclusiones de estudios e investigaciones –1.3-Estudio–; tratan de proyectos e iniciativas sobre alimentación –1.5-Objetivo– que tratan de alcanzar objetivos; abordan formas de ayudar a conseguir perfeccionar el cuerpo –1.4-Bueno–; atienden a aspectos tanto nutricionales de los que dan recomendaciones –1.6-Nutricional– como del sector alimentario –1.7-Alimentario–; y tienen una importancia especial los aspectos de la educación en la alimentación de los niños, tanto en la familia como en la escuela –1.8-Importante y 1.9-Menú–. La última comunidad –1.10-Poner– solo tiene dos palabras.

4.4. La comunidad 2-Comer

Para analizar la segunda comunidad, 2-Comer, menos extensa y con un mayor grado de modularización (vid Tabla 5), lo primero que haremos será modularizarla, hemos detectado 15 subcomunidades; en segundo lugar, determinaremos el grado de importancia de cada una de las subcomunidades reconocidas en el conjunto de la comunidad y del corpus.

El grado de importancia, como ya hemos hecho anteriormente, lo determinaremos a través de su grado de intermediación, por lo que atenderemos al grado de intermediación de la palabra clave de cada una de las comunidades (vid. Tabla 8), aquellas comunidades cuya palabra clave ocupe un puesto muy bajo en el conjunto de la comunidad podemos sospechar que representa un tema local o particular de los textos y no uno general del corpus.

Tabla 8. Subcomunidades de 2-Comer con su palabra clave, su intermediación y el puesto que ocupa en el conjunto de la comunidad a partir de su grado de intermediación.

Subcomunidad	Grado	Intermediación	Tamaño	Puesto
2.1-Comer	159	0,166	151	1
2.2-Comida	125	0,121	119	2
2.3-Dar	96	0,088	83	3
2.4-Ver	85	0,078	95	4
2.5-Llegar	69	0,051	59	7
2.6-Hora	65	0,037	65	9
2.7-Pequeño	52	0,034	30	13
2.8-Momento	55	0,033	43	14
2.9-Sabor	45	0,030	45	17
2.10-Ciencia	42	0,023	43	23
2.11-Tiempo	53	0,022	45	25
2.12-Consistir	19	0,016	47	35
2.13-Mano	31	0,014	27	38
2.14-Blog	11	0,007	12	67
2.15-Combinar	23	0,006	28	70

Vemos como, a partir de la subcomunidad 2.6-Hora, los puestos que ocupan las palabras clave son muy bajos, por lo que consideraremos que esas subcomunidades son locales; además, el bajo grado de intermediación de las comunidades a partir de la 2.2-Comida nos hace sospechar que esas subcomunidades no tienen un gran peso dentro del corpus general.

Vamos ahora a atender a las palabras con grado mayor de 9 dentro de cada una de las subcomunidades que hemos seleccionado:

Comunidad 2.1.-Comer: comer (59), forma (33), manera (28), sano (27), cosa (25), libro (22), casa (19), tiempo (18), preparar (18), natural (17), pensar (17), adelgazar (17), necesitar (16), alimentar (16), autor (16), aprender (15), cambiar (14), trabajar (13), claro (12), dedicar (11), reto (10).

Comunidad 2.2.-Comida: comida (48), producir (22), rápido (14), empezar (13), posibilidad (13), cena (11), casero (11).

Comunidad 2.3-Dar: dar (32), lugar (17).

Comunidad 2.4-Ver: ver (31), gente (18), época (14), decidir (11).

Comunidad 2.5-Llegar: llegar (20), pasar (12).

Comunidad 2.6-Hora: hora (22), medio (15).

Tal como sospechábamos, parece que a partir de la subcomunidad 2.3-Dar, lo que tenemos son comunidades formadas por subtemas individuales unidos por unas pocas palabras que presentan un grado relativamente alto que son las que los mantienen unidos; como muestra podemos ver el gráfico de la comunidad 2.3-Dar (Figura 8) y de la 2.4-Ver (Figura 7).

En ambos grafos se puede ver cómo las palabras clave son casi las únicas que ocupan el centro y de ellas parten los brazos de los subtemas que, entre sí, mantienen pocas o ninguna relación.

Esta situación justifica que podamos quedarnos con las dos primeras subcomunidades como aquellas que representan los temas generales que aparecen en la comunidad 2-Comer. Ambas tratan la forma de comer, comida casera o rápida, y de su preparación; en estos dos ámbitos, podemos destacar palabras como *natural*, *aprender*, *sano*, *cambiar*, *adelgazar* que nos indican el foco de atención en ambos. No parece que se trate de cómo se prepara un plato, sino de cuál debe ser el objetivo de su preparación: la salud.

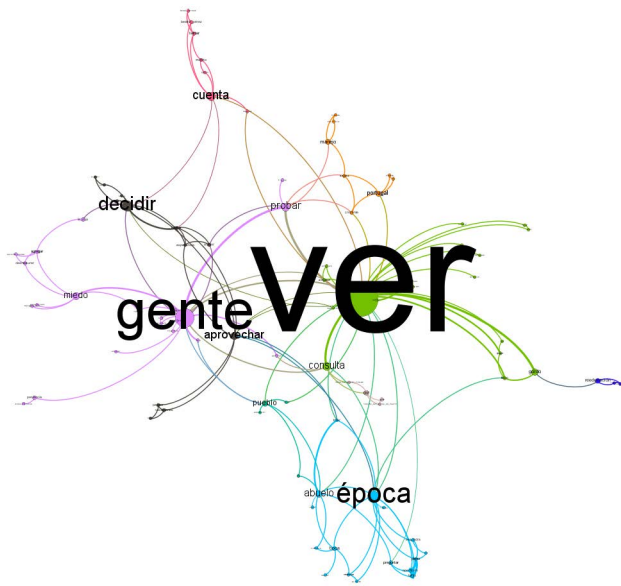


Figura 7. Grafo de la comunidad 2.4-Ver.

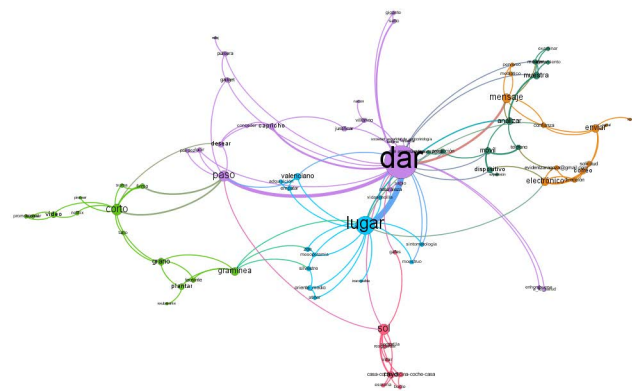


Figura 8. Grafo de la subcomunidad 2.3-Dar.

4.5. Las otras comunidades

El resto de comunidades del corpus, desde la comunidad 3-año a la 10-Dieta, no van a ser modularizadas porque sus subcomunidades suelen tener pocas palabras con importancia dentro del corpus; eso no quiere decir que las comunidades en sí no lo tengan, sino que cada una lo tiene como grupo, siendo sus subgrupos más locales. Algunas de las palabras presentes en estas comunidades tienen un carácter de subforma léxica como “producto”, “persona”, “tema” o “información”, por ejemplo. Otras son terminología especializada pero de uso frecuente en el ámbito del discurso alimentario, como “caloría” o “carbohidrato”. Y también observamos terminología relacional como “mismo”, que aporta coherencia al discurso al vincular distintos elementos en los textos y establecer relaciones entre ellos. Lo interesante es que el análisis semántico de las palabras de cada una de estas comunidades revela la proximidad significativa de los términos y abunda en la construcción canónica de los discursos sobre alimentación en la prensa en función de las temáticas tratadas.

Al tratarlas de forma global hemos escogido como palabras que nos pueden indicar más detalladamente su tema las 15 primeras con el grado más alto:

Comunidad 3-Año: año (137), mayor (100), persona (93), obesidad (90), peso (69), vez (65), sobrepeso (63), nivel (62), población (62), país (53), español (52), España (51), mujer (51), aumentar (50).

Comunidad 4-Fruto: fruto (128), día (81), verdura (81), pescado (73), huevo (72), leche (71), tomar (67), aceite (64), plato (60), arroz (59), ensalada (58), carne (57), aportar (51), mismo (48), caloría (46).

Comunidad 5-Alimento: alimento (147), producto (103), consumo (72), consumir (72), tipo (64), evitar (64), agua (61), bebida (50), alcohol (45), recomendar (43), reducir (43), ingesta (42), determinar (37), mercado (35), refresco (31).

Comunidad 6-Azúcar: grasa (68), proteína (64), cantidad (59), añadir (55), azúcar (52), nutriente (45), vitamina (45), rico (44), gran (43), fibra (42), ácido (37), llevar (36), hierro (36), contener (35), energía (34).

Comunidad 7-Explicar: explicar (58), experto (50), doctor (41), científico (39), presidente (36), director (36), publicar (27), apuntar (24), revista (22), asegurar (18), miembro (18), profesor (18), empresa (16), genético (16), organizar (15).

Comunidad 8-Problema: enfermedad (68), problema (67), riesgo (37), diabetes (34), provocar (30), sufrir (29), deber (27), factor (25), cardiovascular (24), patología (23), aumento (22), causa (20), estrés (18), muerte (18), proteger (17).

Comunidad 9-Salud: salud (57), efecto (42), físico (29), creer (25), estado (22), grande (21), beneficio (21), actividad (20), solo (20), público (20), punto (20), impacto (19), afectar (19), tema (18), suelo (17).

Comunidad 10-Dieta: dieta (105), seguir (31), calidad (30), información (27), grupo (25), hecho (13), adecuar (15), humano (16), advertir (12), mito (18), vegano (13), técnica (13), mediterráneo (23), carbohidrato (12), productor (9).

Estas 8 comunidades completan el panorama y señalan claramente los diferentes enfoques o subtemas que se han tratado de los temas generales; entre ellos, agrupando un poco las comunidades para organizar, podríamos distinguir tres grandes grupos que podemos denominar *salud*, *nutrientes* y *expertos*.

Dentro del que hemos denominado grupo de *salud* se encuadra claramente la *Comunidad 9-Salud*, pero también la *Comunidad 3-Año* que aborda fundamentalmente la obesidad y el sobrepeso con especial atención a la mujer o la *Comunidad 8-Problema*, en la que la alimentación se ve como una causa de los numerosos problemas de salud, enfermedades y patologías.

En el segundo grupo, *nutrientes*, podemos agrupar a las comunidades *4-Fruto*, *5-Alimento*, *6-Azúcar* y *10-Dieta*. En estas comunidades se aborda los productos que se han de consumir y los que se han de evitar para constituir una dieta adecuada; de estos productos o alimentos lo que interesa son los nutrientes que aportan a la dieta, de ahí que aparezcan palabras como grasa, proteína, carbohidrato, vitamina, fibra.

4.6. Discusión general

Llegados a este punto, podemos ahora hacer una valoración general de la temática del corpus de alimentación agrupando y presentando una visión global de los resultados obtenidos. Para comenzar, hay una conclusión que se puede extrapolar de forma clara: la comida ya no es un placer, sino un problema.

Así que el adagio latino “*edamus, bibamus, gaudeamus*” se convierte ahora en una aberración social que debe ser censurada, pues la alimentación sólida o líquida debe estar enfocada con parámetros exclusivamente nutricionales. Los alimentos ya no son fuente de placer, ni la comida un momento de relajación y relación social, sino que son vistos como un producto cuyo objetivo fundamental es la salud. Estos productos, desprovistos de textura y gusto, sirven para mejorar la salud o para solucionar o evitar sus problemas (enfermedades, patologías, obesidad).

Los alimentos se dividen ahora en dos clases, los que se han de evitar y los que se han de recomendar y se clasifican a partir de los beneficios que estos tengan para la salud. La recomendación o no de un alimento se basa no en cuestiones de gusto, sino en sus nutrientes. Los sabores (dulce, amargo, ácido, agridulce...) y las texturas (crujiente, blando, viscoso, harinoso...) son sustituidos por vitamina, proteína, fibra, hierro... De este modo, cuando vayamos a comer, cuando prepares tu comida, no debes primar el paladar, sino la salud; no debes preguntarte cómo está hecho, sino de qué está hecho que es lo que realmente hace a los alimentos buenos o malos.

Como hemos indicado, la divulgación de la alimentación centra ahora su foco en las consecuencias de la comida en la salud y fija su atención en las enfermedades provocadas y favorecidas por esta. Ahora el centro de atención se desplaza a los nutrientes y a sus efectos enfermedades cardiovasculares, diabetes, estrés, cáncer, etcétera, y, dentro de estas patologías, la obesidad es la estrella. En futuras investigaciones se nos presenta el desafío de analizar a través de herramientas como AntConc la frecuencia de coapariciones de verbos como “causar” y términos como “enfermedad” o “muerte” en la totalidad del corpus DAME (Discurso de la Alimentación en los Medios), así como la relación con los alimentos o dietas que se identifican como causas.

La obesidad, no diferenciada muchas veces del sobrepeso, se convierte en el gran anatema, la gran maldición social que, más allá de afectar a una parte importante y preocupante de la sociedad, se extiende a todos los humanos. Todos, de una manera u otra, debemos estar alerta con respecto a nuestra alimentación, padecemos o no exceso de peso.

Debemos estar alerta porque, ahora, la función de la alimentación es mejorar la salud y perfeccionar el cuerpo; no importa la salud de la que se goce, en el fondo, todo el mundo debe llevar una dieta saludable. Esta frase “dieta saludable”, que parece razonable, creemos que conlleva una pequeña trampa.

La dieta no es vista simplemente como el conjunto de alimentos que ingieres, sino como una planificación de esa ingesta, como un régimen alimenticio, de manera que, en el análisis general que hemos hecho se observa, que, al final, todos somos o podemos ser enfermos y todos debemos estar vigilantes y atentos a los productos que comemos.

Ahora bien, dentro de ese todos hay dos colectivos que centran la atención: las mujeres y la infancia. En el segundo grupo, parece claro que se debe a la preocupación, ya expresada por la OMS, sobre el crecimiento del porcentaje de niñas y niños con obesidad o sobrepeso en el mundo; los textos se centran en la dieta que siguen los y las menores tanto en la escuela como en casa, por lo que la educación es un tema esencial en este ámbito.

El motivo de que la infancia sea un centro de atención, como ya hemos dicho, parece claro, los datos son preocupantes, pero las razones para centrarse en las mujeres ya no lo son tantas. La comparación de datos sobre obesidad mundial de hombres y mujeres no muestra grandes diferencias: en 2016 solo hay una diferencia del 1% en cuanto al sobrepeso y un 4% en cuanto a la obesidad (Obesidad y sobrepeso, s. f.) Estas

diferencias no justifican que las mujeres, frente a los hombres, se conviertan en un tema central en torno a la alimentación, la dieta y la obesidad. La razón de su importancia temática hay que buscarla en la proyección de roles de género estereotipados sobre la imagen, la autoimagen y el cuidado personal de las mujeres (Rodríguez Barcia, Varela Suárez y Rifón Sánchez, 2022).

La cobertura del pastel, producto no recomendable por contener grasas y diversos nutrientes diabólicos, de la divulgación mediática de la alimentación se hace con una buena capa de merengue de prestigio científico. Como se ha visto en el análisis, las comunidades en las que aparecen lemas como *experto*, *investigador*, *doctor*, *científico* junto con *revista*, *publicar* o *demostrar* crean la cobertura perfecta para dar a la información un aspecto apetecible a un público no especializado pues aportan el prestigio y el argumento de autoridad que la audiencia necesita para considerar confiables estas afirmaciones. Esperamos no indigestarnos.

5. Conclusiones

Para concluir, las medidas de modularidad de las comunidades, el grado de intermediación y el grado de la teoría de grafos o redes, han determinado cómo el discurso divulgativo mediático sobre la alimentación se centra, fundamentalmente, en los efectos de esta sobre la salud, con la obesidad como tema estrella. Los colectivos más tratados son la infancia y las mujeres, para lo cual se divulgan propiedades y efectos de los nutrientes que componen los alimentos, lo cual podemos entender como un proceso creciente de *nutricialización* de la comida. Para ello, se reelaboran y se seleccionan estratégicamente argumentos científicos.

Contribución de autoría CREDIT

Antonio Rifón Sánchez (A. R. S.), Susana Rodríguez Barcia (S. R. B.) y Ana Varela Suárez (A. V. S.). La aportación realizada por cada uno de los autores del artículo es la que se indica a continuación: Conceptualización: A. R. S., S. R. B., A. V. S.; Análisis: A. R. S., S. R. B., A. V. S.; Metodología: A. R. S.; Supervisión: S. R. B., A. V. S.; Redacción del borrador original: A. R. S., S. R. B., A. V. S.; Redacción de la revisión y edición final: S. R. B. y A. V. S.

Referencias bibliográficas

- AIMC. (2018). *Resumen General de Resultados EGM (2017-2018)*. <https://www.aimc.es/a1mc-c0nt3nt/uploads/2018/06/resumegm218.pdf>
- Amancio, Diego Raphael (2015). A Complex Network Approach to Stylometry. *PLOS ONE*, 10(8), e0136076. <https://doi.org/10.1371/journal.pone.0136076>
- Annandale, Ellen y Clark, Judith (2000). Gender, postmodernism and health. En J. Gabe y M. Calnan, *Health, medicine and society: Key theories, future agendas*, (pp. 51–64).
- Antiqueira, Lucas et al. (2007). Strong correlations between text quality and complex networks features. *Physica A: Statistical Mechanics and its Applications*, 373, 811–820. <https://doi.org/10.1016/j.physa.2006.06.002>
- Antiqueira, Lucas et al. (2005). Modelando textos como redes complejas. *Anais do III Workshop em Tecnologia da Informação e da Linguagem Humana*, 22–26. <http://nilc.icmc.usp.br/til/til2005/arq0054.pdf>
- Bastian, Mathieu, Heymann, Sébastien y Jacomy, Mathieu (2009). Gephi: An open source software for exploring and manipulating networks. *ICWSM*, 8, 361–362. <https://doi.org/10.1609/icwsml.v3i1.13937>
- Boudin, Florian (2018). Unsupervised Keyphrase Extraction with Multipartite Graphs. *arXiv preprint arXiv:1803.08721*. <https://arxiv.org/pdf/1803.08721.pdf>
- Bougouin, Adrien, Boudin, Florian y Daille, Béatrice (2013). Topicrank: Graph-based topic ranking for keyphrase extraction. *International Joint Conference on Natural Language Processing (IJCNLP)*, 543–551. <https://hal.archives-ouvertes.fr/hal-00917969/document>
- Brandes, Ulrik (2001). A faster algorithm for betweenness centrality*. *Journal of Mathematical Sociology*, 25(2), 163–177. <https://doi.org/10.1080/0022250X.2001.9990249>
- Carreras, Xavier et al. (2004). FreeLing: An Open-Source Suite of Language Analyzers. *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC'04)* 239–242. https://www.researchgate.net/profile/Muntsia_Padro/publication/228976391_Freeling_An_open-source_suite_of_language_analyzers/links/02bfe50fd836ec3df3000000.pdf
- Cassany, Daniel, López, Carmen y Martí, Jaume (2000). Divulgación del discurso científico: La transformación de redes conceptuales. Hipótesis, modelo y estrategias. *Discurso y sociedad*, 2(2), 73–103.
- Cassany, Daniel y Martí, Jaume (1998). Estrategias divulgativas del concepto príon. *Quark*, 12, 58–66.
- Cook, Guy (2004). *Genetically modified language: The discourse of arguments for GM crops and food*. Routledge.
- Cook, Guy (2010). Sweet talking: Food, language, and democracy. *Language teaching*, 43(2), 168–181. <https://doi.org/10.1017/S0261444809990140>
- de Benito Morán, Cristina (2016). *La alimentación como práctica política cotidiana: Análisis antropológico social de los grupos de consumo ecológico* [PhD Thesis]. Universidad Autónoma de Madrid.
- Díaz Rojo, José Antonio y Morant Marco, Ricard (2007). El discurso crítico contra la "tiranía" del culto al cuerpo. *Tonos digital*, 14.

- Díaz Rojo, José Antonio, Morant Marco, Ricard y Westall, Debra (2005). El sanismo lingüístico: Recursos retóricos en la publicidad y etiquetado de los alimentos. *Revista de investigación lingüística*, 8, 35–52.
- Ferreira da Silva, Luisa (2013). Food and culture: Change and continuity/A questao cultural na alimentacao: mudancas e permanencias. *Demetra: Food, Nutrition & Health*, 8(2 SUP), 247–265. <https://doi.org/10.12957/demetra.2013.6904>
- Fortunato, Santo (2010). Community detection in graphs. *Physics reports*, 486(3), 75–174. <https://doi.org/10.1016/j.physrep.2009.11.002>.
- Garg, Muskan y Kumar, Mukesh (2018). Identifying influential segments from word co-occurrence networks using AHP. *Cognitive Systems Research*, 47, 28–41. <https://doi.org/10.1016/j.cogsys.2017.07.003>.
- Gouveia, Carlos y Alexandre, Marta Filipe (2009). The Arrogant Scientist and the Ignorant Citizen: A Critical Discourse Analysis of the Discourse of Scientis. En A. Pinheiro de Sousa (Ed.), *So long lives this, and this gives life to thee: Homenagem a Maria Helena de Paiva Correia* (pp. 177-190). Edições Colibri.
- Gracia Arnaiz, Mabel (2007). Comer bien, comer mal: La medicalización del comportamiento alimentario. *Salud pública de México*, 49(3), 236–242.
- Gracia Arnaiz, Mabel y Comelles, Josep (Eds.). (2007). *No comerás: Narrativas sobre comida, cuerpo y género en el nuevo milenio*. Icaria Editorial.
- Jarosz, Lucy (2011). Defining world hunger: Scale and neoliberal ideology in international food security policy discourse. *Food, culture & society*, 14(1), 117–139. <https://doi.org/10.2752/175174411X12810842291308>
- Lahiri, Shibamouli (2013). Complexity of word collocation networks: A preliminary structural analysis. *arXiv preprint arXiv:1310.5111*.
- López Torres, Leyna Priscila y Contreras Hernández, Jesús (2020). Percepciones sobre la alimentación saludable y sus implicaciones en la cocina doméstica en un grupo de mujeres de Cataluña, España. *Perspectivas En Nutrición Humana*, 22(1). Recuperado a partir de <https://revistas.udea.edu.co/index.php/nutricion/article/view/338415>.
- Lupton, Deborah (2000). Food, risk and subjectivity. En J. Gabe y M. Calnan *Health, Medicine, and Society: Key Theories, Future Agendas London: Routledge*, (pp. 205–18).
- Marauri Castillo, Íñigo et. al. (2020). Instagram y la información sobre nutrición en España. En K. Meso Ayerdi, S. Peña Fernández y A. Larrondo Ureta (ed.) *Nuevos actores y estrategias en el espacio mediático compartido*. Universidad del País Vasco, (pp. 205–223).
- MacMillan, Tom y Dowler, Elizabeth (2012). Just and sustainable? Examining the rhetoric and potential realities of UK food security. *Journal of agricultural and environmental ethics*, 25(2), 181–204.
- Mihalcea, Rada y Radev, Dragomir (2011). *Graph-based natural language processing and information retrieval*. Cambridge University Press.
- Newman, Mark (2010). *Networks: An introduction*. Oxford University Press.
- Padró, Lluís y Stanilovsky, Evgeny (2012). Freeling 3.0: Towards wider multilinguality. *LREC2012*. <https://upcommons.upc.edu/handle/2117/15986>.
- Paranyushkin, Dmitry (2011). Identifying the pathways for meaning circulation using text network analysis. *Nodus Labs, Berlin*. <http://noduslabs.com/publications/Pathways-Meaning-Text-Network-Analysis.pdf>.
- Reiss, Julian y Sprenger, Jan (2017). Scientific Objectivity. En E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2017). Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/win2017/entries/scientific-objectivity/>.
- Rifón, Antonio (2020). Progresión temática y cohesión textual a través de grafos de coocurrencias. *Círculo de Lingüística Aplicada a la Comunicación*, 82, 193–208. <https://doi.org/10.5209/clac.68972>.
- Rodríguez Barcia, Susana; Varela Suárez, Ana y Rifón Sánchez, Antonio (2022). Visibilidad y construcción discursiva de la mujer en las publicaciones sobre alimentación de la prensa española. *Discurso & Sociedad* 16(2), 383–419.
- Zhou, Zhi et al. (2013). Research on weighted complex network based keywords extraction. *Workshop on Chinese Lexical Semantics*, 442–452. https://doi.org/10.1007/978-3-642-45185-0_47.